Functional analytic approaches to some stochastic optimization problems

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

doctor rerum naturalium
(Dr. rer. nat.)
im Fach Mathematik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von
Herrn Julio Daniel Backhoff

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Jan-Hendrik Olbertz

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät:
Prof. Dr. Elmar Kulke

Gutachter:
1. Prof. Dr. Ulrich Horst
2. Prof. Dr. Joseph Frédéric Bonnans
3. Prof. Dr. Michael Kupper

eingereicht am: 16.09.2014
Tag der Verteidigung: 30.01.2015

Abstract

In this thesis we shall deal with utility maximization and stochastic optimal control problems through several modern points of view. We shall be interested in understanding how such problems behave under parameter uncertainty under respectively the robustness paradigm and the first order sensitivity approach. Afterwards, we shall leave the single-agent world and tackle an instance of a two-agent problem where the first one delegates his/her investments to the second through a contract.

In the first place, we consider the robust utility maximization problem in continuous time financial market models, where we formulate conditions for the partial and full solvability of the problem without assuming weak compactness of the densities of the so-called uncertainty set, which is a set of measures upon which the utility maximizing agent wants to perform robust investments. These conditions are stated in terms of functional spaces that arise naturally from the formulation of the problem. For general markets, we show that the relevant space is a certain Modular space, through which we can prove a minimax equality and the existence of optimal strategies. In complete markets the relevant space is an Orlicz space, and upon granting its reflexivity under verifiable conditions on the utility function, we obtain in addition the existence of a worst-case measure in the uncertainty set. We moreover characterize the latter in terms of the solution to a certain bi-dual problem which can in practical cases be simpler to solve.

Secondly we turn our attention to continuous-time stochastic optimal control, where we provide a first order sensitivity analysis to some parameterized variants of such problems. The main tool here is the one-to-one correspondence, which we rigorously prove, between the adjoint states appearing in a weak form of the stochastic Pontryagin principle and the Lagrange multipliers associated to the controlled equation when viewed as a functional constraint on a space of processes. The sensitivity analysis is then deployed in its full strength in the case of convex problems and additive perturbations as well as in specific mean-variance or linear-quadratic problems and multiplicative perturbations.

In a final part, we proceed to Principal-Agent problems in discrete time. Here we apply in the greatest possible generality the tools from conditional analysis to the case of linear contracts and show that most results known in the literature for very specific instances of the problem carry on to a much larger family of utility functions and probabilistic settings. In particular, the existence of a first-best optimal contract and its implementability by the Agent is recovered.

Zusammenfassung

In dieser Arbeit beschäftigen wir uns mit Nutzenoptimierungs- und stochastischen Kontrollproblemen unter mehreren modernen Gesichtspunkten. Wir untersuchen die Parameterunsicherheit solcher Probleme, einerseits im Sinne des Robustheitsparadigma und andererseits bezüglich des ersten Ordnungssensitivitätsansatzes. Neben der Betrachtung dieser Einagentenproblemen widmen wir uns auch einem Zweiagentenproblem, bei dem der eine Agent dem anderen das Management seines Portfolios vertraglich überträgt.

Wir betrachten das robuste Nutzenoptimierungsproblem in zeitstetigen Finanzmarktmodellen, wobei wir hinreichende Bedingungen für die partielle und vollständige Lösbarkeit des Problems formulieren, ohne jegliche schwache Kompaktheit der sogenannten Unsicherheitsmenge zu fordern, welche die Maße enthält, auf die der Optimierer robustifiziert. Unsere Bedingungen sind über gewisse Funktionenräume beschrieben, die sich in einer natürlichen Weise aus der Formulierung des Optimierungsproblems ergeben. Für allgemeine Märkte zeigen wir, dass der passende Raum ein bestimmter Modularraum ist, mittels dem wir eine Min-Max-Gleichung und die Existenz von optimalen Strategien beweisen können. In vollständigen Märkten ist der relevante Raum ein Orlicz-Raum, und nachdem man seine Reflexivität mithilfe verifizierbarer Bedingungen überprüft hat, erhält man zusätzlich die Existenz sogenannter Worst-Case-Maße innerhalb der Unsicherheitsmenge. Weiterhin charakterisieren wir diese Maße anhand der Lösung eines bestimmten bi-dualen Problems, welches in spezifischen Fällen einfacher zu lösen ist.

Für die Parameterabhängigkeit stochastischer Kontrollprobleme in stetiger Zeit entwickeln wir einen Sensitivitätsansatz erster Ordnung. Das Kernargument ist hier die Korrespondenz, die wir rigoros beweisen, zwischen dem adjungierten Zustand zur schwachen Formulierung des Pontryaginschen Prinzips und den Lagrange-Multiplikatoren, die der Kontrollgleichung assoziiert werden, wenn man sie als eine funktionale Bedingung auf einem Raum von Prozessen betrachtet. Der Sensitivitätsansatz wird dann in voller Stärke auf konvexe Probleme mit additiver Störung, sowie für spezifische Mean-Variance- und linear-quadratische Probleme mit multiplikativer Störung angewendet.

Das Prinzipal-Agent-Problem formulieren wir in diskreter Zeit. Wir wenden in größter Verallgemeinerung die Methoden der bedingten Analysis auf den Fall linearer Verträge an und zeigen, dass sich die Mehrheit der in der Literatur unter sehr spezifischen Annahmen bekannten Ergebnisse auf eine deutlich umfassenderer Klasse von Nutzenfunktionen und probabilistischer Kontexte verallgemeinern lässt. Insbesondere erhalten wir weiterhin die Existenz eines first-best-optimalen Vertrags und dessen Implementierbarkeit durch den Agenten.

Contents

1 Introduction

In this thesis we shall deal with utility maximization and stochastic optimal control problems through several modern perspectives and mathematical lenses. One aspect that will interest us is the question of dependence of a stochastic optimization problem with respect to its defining model parameters. There are at least two ways of assessing this, which in a sense lie at opposite poles; on the one hand the robust approach whereby a worst-case point of view with respect to parameter uncertainty is implemented and on the other hand the sensitivity approach in which case one tries to understand the infinitesimal behaviour of an optimization problem's value as its parameters are slightly varied. The first approach is carried out in Chapter 2 where the issue of robustness in the problem of maximization of expected utility from terminal wealth in continuous time is tackled through convex duality and functional analytical arguments. Then in Chapter 3 we deal with the second approach applied to stochastic optimal control problems, showing how the solution to the Backward Stochastic Differential Equation in Pontryagin's Principle encodes the marginal dependence of the optimal value of the problem with respect to say the drift and the volatility parameters in the concrete model. A further question of interest in this thesis is the more modern and realistic situation in which the utility maximizing agent delegates to a third party the management of his/her wealth through an incentive providing contract. In Chapter 4 we employ the recently developed theory of conditional analysis to deal with the mentioned dynamic agency problem in discrete time and under linear contracts for very general utility functions and probabilistic frameworks.

The problem of expected utility maximization in continuous time models of financial markets has been thoroughly researched in the last decades. For a complete solution of the very well understood frictionless case (and without consumption) we refer to Kramkov and Schachermayer [1999], which is the culmination of a long line of related works, and the references therein, where the authors use convex duality methods as well as some pseudo-notions of compactness on the non-locally convex space of measurable functions to fully characterize the solution of the problem, even in the case of incomplete markets.

However, in a standard utility maximization problem one is forced to choose (or say fix) a probability measure $\mathbb{P}$ under which the random objects in the model shall evolve. It goes without a saying that in practical terms it is next to impossible to, with complete accuracy, compute this real-world measure. For instance, any statistical method shall only sign out a region of confidence for it, and not a single one. Therefore one is quickly led to consider utility maximization under families of possible measures (we refer to this as the *uncertainty set* and denote it $\mathcal{Q}$) rather than over a unique a priori one;

see Gilboa and Schmeidler [1989] for more on this idea. A commonly adopted (though conservative) point of view is to look for strategies that are optimal in the worst possible sense:

$$\text{maximize } \inf_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}}\left[\text{Utility}(X)\right] \text{ over all admissible terminal wealths } X.$$

We will also consider in Chapter 2 such a point of view and, as usual in the literature, we shall refer to this stochastic optimization problem as the robust variant of the (standard, non-robust) utility maximization one.

In Quenez [2004], Gundel [2005], Schied and Wu [2005], Schied [2005], Föllmer and Gundel [2006], to name a few, the problem of robust utility maximization from terminal wealth is solved in a way that greatly recovers the results known for the non-robust situation. In more concrete terms; under the assumption that the uncertainty set enjoys some sort of compactness and its measures are dominated by a single, reference one, the authors successfully apply convex-duality arguments and deliver attainability of the problem (as well as its dual, conjugate problem) and even the existence of what may be called a *worst-case measure*, which is a measure in the given family for which the optimal utility is as low as it gets. We should say at this point that in the presence of consumption the robust instance of the problem has also been considered in e.g. Burgert and Rüschendorf [2005] and Wittmüss [2008], and that in general convex analysis is not the only way to tackle these problems: see Hernández-Hernández and Schied [2007] for a stochastic control approach (via PDEs), as well as Bordigoni et al. [2007] and the references therein for an approach using BSDEs. However, the assumption of compactness on the family of possible measures seems prevalent in the literature, whatever the approach. Moreover whereas some sort of explicit characterization for the optimal wealth (strategy) for the problem is typically deduced, very little is said about the worst-case measure in concrete terms, beyond very specific instances of the problem.

The usual actual assumption of compactness of the uncertainty set is specifically that the densities of the measures therein with respect to a fix reference one $\mathbb{P}$ form a uniformly integrable set. Looking at an extremely simple instance of the problem (see Examples 2.2.1 and 2.2.3) suggests that both this compactness assumption and the lack of a systematic characterization for the worst-case measure could be tackled with general techniques and tools of convex duality. For instance consider that the family of measures came out of the intersection of a hyperplane (in the space of signed measures) with the set of probability measures. Then the densities of this family are certainly not expected to enjoy any compactness property a priori (we will later provide an explicit example of this situation), as hyperplanes are quite unbounded in most senses. However, when seen as an infinite-dimensional optimization problem, the dual of the robust utility maximization problem turns out to be, in this particular case, what is called a *convex problem*: to minimize a convex functional under linear-convex constraints. Therefore, there is every reason to believe that an a priori compactness requirement of the feasible set of measures could be relaxed under some wider structural assumptions on the problem, and that a full characterization of the solution should be available, in particular meaning an explicit expression for the worst-case measure. This is also suggested by relatively recent developments on general entropy minimization problems (see e.g. Léonard [2008], Léonard [2010] and references therein), which deal with such situations.

In this work we shall only consider the case of utility functions on the positive half line. Our approach will consist on finding an appropriate Banach space where the potential worst-case measures should a fortiori lie. This space will turn out to be a *Modular space* (see Musielak [1983]) and its norm will be closely related to the elements of the optimization problems at hand (more concretely, to the dual problem related to the Legendre transform of the utility function). The crucial argument, and the point where most mathematical difficulties arise, is to prove under verifiable conditions on the utility function and the market that the image through the utility function of all possible terminal wealths is a bounded set $\mathcal{K}$ contained in the norm-dual of the mentioned modular space, meaning that the robust utility maximization reduces to solving:

$$\text{maximize } \inf_{Q \in \mathcal{Q}} \mathbb{E}^{\mathbb{P}} \left[\frac{d\mathbb{Q}}{d\mathbb{P}} K \right] \text{ over } K \in \mathcal{K},$$

providing us with the needed compactness which we use in Theorems 2.2.3 and 2.5.1 to prove the usual minimax equality as well as the existence of optimal wealth processes even when $\mathcal{Q}$ is quite arbitrary. We thus recover some of the results in Schied and Wu [2005], Föllmer and Gundel [2006] et al. under the less stringent assumption that the densities of the uncertainty set be closed with respect to the modular space topology. We also stress that we divert from the usual paradigm of finding a worst-case measure first and then the optimal wealth (thus only defined up to the support of such measure); indeed, we find an optimal wealth defined over the whole support of the reference probability measure even if a worst-case measure does not exist. Let us mention that we envision that the compactness of the image by the utility function of the final admissible wealths should become a fruitful argument for problems beyond robust utility maximization, and already in the non-robust case it sheds new insight into the subject (see Proposition 2.5.6).

When we set ourselves to recover or sharpen those results in Schied and Wu [2005], Föllmer and Gundel [2006] et al. not covered by the aforementioned approach, for instance the existence of a worst-case measure or the characterization of the optimal wealths, we realize that reflexivity of the Modular space is a sufficient means to doing this. In this respect we prove, modulo some pathologies on the filtered probability space, that our Modular spaces are unfortunately never reflexive for strict incomplete markets; this is the content of Theorem 2.5.2 and the remarks thereafter. On the positive side, when we specialize our analysis to complete markets, our Modular spaces become Orlicz-Musielak spaces (a generalization of L^p spaces) and we can provide easily verifiable conditions under which they do become reflexive, from which most of the results in the literature are recovered even under the less stringent hypothesis of closedness of the densities of the uncertainty set with respect to the given Orlicz-Musielak topology, as we prove in Theorem 2.2.5. We should stress that Orlicz spaces are of course known about in Mathematical Finance, for instance through the articles Cheridito and Li [2009] regarding risk measures or Biagini and Frittelli [2008] regarding utility maximization and admissibility of trading strategies. In Föllmer and Gundel [2006] the authors use the concept of f-divergences (see also Goll and Rüschendorf [2001] and the references given) to study the robust utility maximization problem and a Vallee Poussin criterion which is connected to a certain Orlicz space. The approach to the robust problem we shall develop here is different and makes a more systematic use of such type of spaces and its generalizations (Modular spaces).

In the complete market case, this Orlicz space formulation of the (dual) problem will also allow us to describe the worst-case measure. More precisely, by writing the set of possible models in terms of a potentially infinite system of linear constraints, we will be able to give an explicit characterization of the worst-case measures in a much more general setting than it is available in the literature, in the reflexive case. To that end we will adapt to the present setting in Theorems 2.2.6 and 2.4.2 general entropy minimization techniques developed in Léonard [2010] and Léonard [2008] (see the references therein as well) in order to characterize the worst-case measure $\hat{\mathbb{Q}} \in \mathcal{Q}$ in terms of the solution g of a related abstract convex optimization problem (which we may call with some abuse the *bi-dual* problem):

$$\hat{Q} \propto \frac{\mathrm{d}[U^{-1}]}{\mathrm{d}x} \left(\text{linear functional}(g)\right).$$

The so-called bi-dual problem may in many practical situations be easier to solve than the original problem (for instance, it may be finite-dimensional). Let us point out that for incomplete markets, a better understanding of the relevant Modular spaces should enable the use of the general entropy minimization techniques developed in Léonard [2008] in order to characterize the attainability of some extension of the dual problem. In the complete case, it is precisely reflexivity that permits to avoid such an extension.

We close Chapter 2 with a more exploratory discussion regarding the potential of the approach in the case when there is no reference measure dominating the uncertainty set $\mathcal{Q}$. Such a setting has already been studied in Denis and Kervarec [2013] in the context of robust utility maximization under the assumption that $\mathcal{Q}$ is weakly compact as a set of measures (i.e. tight), and the main motivation comes from volatility uncertainty or ambiguity. We skip the delicate issue of the definition of stochastic integrals under infinite possibly singular measures, or their aggregation into a single universally measurable process, and instead focus on finding the candidate Modular spaces of the problem and proving some preliminary results suggesting that the approach might render positive insight into the case when the uncertainty set is not weakly compact any-more.

In Chapter 3 we turn our attention to the issue of first-order sensitivity in stochastic optimal control problems. Our starting point is a reinterpretation of one of the most important results in stochastic optimal control theory: the Pontryagin Principle. Introduced and refined by Kushner [1965], Bismut [1976a], Haussmann [1986], Bensoussan [1983] and Peng [1990] among others (see [Yong and Zhou, 1999, Chapter 3, Section 7] for a historical account), in its simplest form it states that almost surely the optimal control minimizes an associated *Hamiltonian*. This Hamiltonian depends on the optimal state and an *adjoint pair*, which solves an associated Backward Stochastic Differential Equation (BSDE for short). Roughly speaking, the mentioned necessary condition appears as one perturbs the optimal control and analyzes up to first order (or second-order, if the volatility term is controlled and the set of admissible controls is non-convex) the impact of such perturbation on the cost function. A natural question that arises is whether by regarding the stochastic optimal control problem as an infinite dimensional optimization problem in an appropriate functional setting, the usual machinery of optimization theory yields an interpretation of the aforementioned adjoint states. From this perspective it is conceivable that fundamental tools such as convex-duality, Lagrange multipliers and non-smooth analysis (to name a few) may shed a different light and provide new interpretations into the field of stochastic optimal control.

The idea of dealing with stochastic optimal control problems from the point of view of abstract optimization theory is not new. In a remarkable article Bismut [1973], the author extends to the stochastic case the results of Rockafellar [1970a] obtained in the deterministic framework. For convex problems, he proves essentially that the solutions of the original optimization problem and its dual, in the sense of convex analysis, must fulfil the conditions appearing in Pontryagin's Principle. In the non-convex case, a very interesting analysis is performed in Loewen [1987] where the author uses non-smooth analysis techniques to tackle the case of a non-linear controlled Stochastic Differential Equation (SDE for short)[1].

In Chapter 3 we develop a rigorous functional framework under which the Lagrangian approach to stochastic optimal control becomes fruitful. As a matter of fact, we relate the adjoint states appearing in the Pontryagin principle with the Lagrange multipliers of the associate optimization problem, thus extending the results of Bismut [1973] in the convex case, by using a different method. In several interesting cases, this result allows us to perform a first order sensitivity analysis of the value function, under random functional perturbations of the dynamics. To the best of our knowledge, this type of sensitivity results had been obtained for finite dimensional perturbations of the initial condition only (see the works Loewen [1987], Zhou [1990, 1991]). We restrict ourselves to a finite-horizon, brownian setting, yet consider the case of non-linear controlled SDEs with random coefficients and the control being present both in the drift and diffusion parts, pointwise convex constraints on the controls, and finite dimensional constraints of expectation-type on the final state. In mathematical language, we deal with problems

[1] More recently, in e.g. Cheng and Yan [2012], Kosmol and Pavon [2001], the Lagrange multiplier technique has been applied formally in order to derive optimality conditions. However, no connexions with Pontryagin's principle are analyzed.

of the form:

$$\left.\begin{aligned}
&\inf_{(x,u)} \mathbb{E}\left[\int_0^T \ell(\omega,t,x(t),u(t))\mathrm{d}t + \Phi(\omega,x(T))\right] \\
&\text{s.t.} \quad x(t) = x_0 + \textstyle\int_0^t f(s,x(s),u(s))\mathrm{d}s + \int_0^t \sigma(s,x(s),u(s))\mathrm{d}W(s),\ \forall\, t \in [0,T[, \\
&\qquad\ \mathbb{E}\left(\Phi_E(x(T))\right) = 0, \quad \mathbb{E}\left(\Phi_I(x(T))\right) \leq 0, \quad u(\omega,t) \in U \ \text{ a.s.}
\end{aligned}\right\} \tag{CP}$$

where ℓ, Φ, f, σ, x_0, Φ_E, Φ_I are the data of the problem, which can be random, satisfying some natural requirements detailed in Section 3.4, and $U \subseteq \mathbb{R}^m$ is a convex set. Under standard assumptions, we have that for every square integrable and progressively measurable control u, there exists a unique solution $x[u]$ of the SDE in (CP). In this sense, problem (CP) can be reformulated in terms of u only and the SDE constraint can be eliminated. However, we have chosen to work with the pair (x,u) and keep the SDE constraint in order to associate to it a Lagrange multiplier, in view of the important consequences of this approach in the sensitivity analysis of the optimal cost of (CP) (see Section 3.5).

By defining a Hilbert space topology on a certain space of Itô processes, we naturally deduce that whenever the Lagrange multipliers associated to the SDE constraint in (CP) exist they must be Itô processes themselves. With this methodology we can prove a one-to-one simple relationship between the aforementioned Lagrange multipliers and the adjoint states appearing in a weak form of Pontryagin's principle. More concretely, we say that (p,q) is a *weak-Pontryagin multiplier* at a solution (x,u) if the same conditions appearing in the usual Pontryagin principle hold true (see [Peng, 1990, Theorem 3]), except for the condition of minimization of the Hamiltonian which is replaced by the weaker statement corresponding to its first order optimality condition (see Section 3.4.1 for a detailed exposition). Thus, it is easily seen that every adjoint pair appearing in the usual Pontryagin principle is a weak-Pontryagin multiplier. In Theorem 3.4.2 we prove that given a weak-Pontryagin multiplier (p,q), the process

$$\lambda(\cdot) := p(0) + \int_0^{\cdot} p(s)\mathrm{d}s + \int_0^{\cdot} q(s)\mathrm{d}W(s), \tag{1.0.1}$$

is a Lagrange multiplier associated to the SDE constraint in (CP). Conversely, every Lagrange multiplier $\lambda(\cdot) = \lambda_0 + \int_0^{\cdot} \lambda_1(s)\mathrm{d}s + \int_0^{\cdot} \lambda_2(s)\mathrm{d}W(s)$, associated to this constraint, satisfies that $\lambda_0 = \lambda_1(0)$ and (λ_1,λ_2) is a weak-Pontryagin multiplier. Let us stress that the main difficulty of this results lies in first having identified the proper Hilbertian topology useful for our problem and then making a link between certain adjoint operators on Itô processes and linear BSDEs. The latter point is a generalization of e.g. [Yong and Zhou, 1999, Chapter 7, Section 2]. What is more, in the case of convex costs and linear dynamics we derive in Theorem 3.5.1 the existence of Lagrange multipliers and hence the Pontryagin principle, by solely invoking the theory of Lagrange multipliers in Banach spaces (see e.g. Bonnans and Shapiro [1998, 2000] for a survey). Even if this type of arguments can be extended to the case of non-convex costs (see

Remark 3.5.1(iv)), at the present time we do not know if it is possible by the latter theory to prove Pontryagin's principle in the case of non-linear dynamics.

One advantage of identifying the Lagrange multipliers of an optimization problem is that, under some precise conditions, these multipliers allow to perform a first-order sensitivity analysis of the value function as a function of the problem parameters. In a nutshell, if the optimization problem at hand is convex (this is the case of convex costs and linear equality constraints) or smooth and stable with respect to parameter perturbations (e.g. if the optimizers converge as we vary the parameters, and the functions involved are at least continuously differentiable) then the sensitivity of the value function in terms of the perturbation is related to the derivative of the Lagrangian with respect to the parameters taken in the perturbation direction (see e.g. [Bonnans and Shapiro, 2000, Section 4.3]).

Using the identification of Lagrange and weak-Pontryagin multipliers we establish in Section 3.5 our main results. In a first part we rely on classical duality theory for convex problems (see e.g. Rockafellar [1974]) and we prove in Theorem 3.5.1, for example, that for stochastic optimal control problems with convex costs and linear dynamics, an additive (random, time-dependent) perturbation $(\Delta f, \Delta\sigma)$ to the drift and diffusion parts of the controlled SDE changes the value function (up to first order) by exactly

$$\mathbb{E}\left(\int_0^T p(t)^\top \Delta f(t)\mathrm{d}t\right) + \mathbb{E}\left(\int_0^T \operatorname{tr}\left[q^\top(t)\Delta\sigma(t)\right]\mathrm{d}t\right),$$

where (p, q) is (in this case) the unique adjoint state appearing in the Pontryagin's principle. A simple corollary of this is that if one perturbs a deterministic optimal control problem by a small (brownian) noise term, the value function remains unaltered up to first-order, as was observed in Loewen [1987] by other methods. Then in Theorem 3.5.2 we provide a version of the previous result when final constraints are considered. We remark that in this case, due to the possible non-uniqueness of the Lagrange multipliers, the directional derivative is not necessarily a linear function of the perturbations. Despite that at the present point we cannot extend the previous sensitivity analysis to general non-convex problems, we do tackle in a second part some cases of non-additive parameter perturbations of convex stochastic optimal control problems. This is an important improvement from what was outlined in the previous paragraph, as in practice parameter error/inaccuracy can propagate in very complicated fashions if for instance this error is amplified by the decision (control) variable. This is the setting we face in two benchmark examples we deal with in this chapter; the stochastic Linear-Quadratic (LQ) control problem and the Mean-Variance portfolio selection problem, which is an LQ problem with a constraint on the expected value of the final state. In these problems, it is natural to consider perturbations of the matrices appearing in the dynamics that multiply either the state or the control. We should underline that for these types of perturbations, classical arguments based on convex analysis as in Rockafellar [1974] are not applicable and more recent results on perturbation analysis have to be invoked (see Bonnans and Shapiro [1998, 2000]). The main tool here is again the identification in

Theorem 3.4.2 and the stability result in Proposition 3.5.1 regarding a weak continuity property for the solutions of linear SDE and BSDE in terms of the parameters, which in both mentioned examples allows us to prove the convergence of the solutions of the perturbed problems.

As suggested by their name, in a stochastic LQ problem one seeks to minimize a quadratic functional of the state and control variables, which are related through a linear SDE. Such problems are to be found everywhere in engineering and economics and we refer the reader to Bismut [1976b], Chen and Yong [2001], Tang [2003], Yong and Zhou [1999] and the references therein for an exposition of the theory. Our main results here are a strong stability property for the solutions of parameterized unconstrained convex LQ problems (see Proposition 3.5.2) and Theorem 3.5.3, where we provide a complete sensitivity analysis for the value function in terms of the parameters. More precisely, we prove that the optimal cost depends in a continuously differentiable manner on the various parameters and we give explicit expressions for the associated derivatives. From the practical point of view, this result may have interesting applications. As matter of fact, recall that the resolution of deterministic LQ problems can be achieved by solving an associated deterministic backward Riccati differential equation. The analogous result holds true in the stochastic framework, see e.g. Tang [2003], but in that case the Riccati equation is a highly nonlinear BSDE. Therefore, for small random perturbations of the matrices of a deterministic LQ problem, it seems reasonable to approximate the value function of the perturbed problem as the value of the deterministic one plus a first order term, which can be calculated in terms of the solution of the deterministic Riccati equation (see Remark 3.5.3(i)).

In the classical Mean-Variance portfolio selection problem, one seeks to find the portfolio rendering the least variance of the terminal wealth with a guaranteed fixed expected return. This is a very central topic in finance and economics, and we refer the reader to Li and Zhou [2000] (random coefficients), Framstad et al. [2004] (case with jumps), among others, for a modern point of view. As for the general LQ case, our major contributions here are Proposition 3.5.3, dealing with the stability analysis for the optimal solutions in terms of the perturbation parameters (the initial capital, deterministic interest/saving rates, the desired return, the drift and the diffusion coefficients) and Theorem 3.5.4, where we prove that the optimal cost is continuously differentiable with respect to those perturbations.

To the best of our knowledge the aforementioned results for the LQ and mean variance problems, regarding the strong stability of the minimizers, the C^1-differentiability of the value functions and the computation of the derivatives for general random perturbations of the dynamics are novel in the literature and certainly we cannot envision at the time any alternative approach yielding similar results/statements.

We close Chapter 3 with a more exploratory sensitivity analysis of the non-robust utility maximization problem, thus providing a bridge to Chapter 2 and yielding a “dual” point of view towards model uncertainty, by looking at the first-order effect of slight misspecification of parameters instead of robustifying on them. We discuss two possible formulations of the problem, and realize that in one of them (which we refer to as the weak one) the parameters to be varied do not enter in the state constraint so typical of such financial problems. We can thus endeavour for this formulation a sensitivity analysis by mixing ideas and result from both referred chapters, the caveat being that often we assume for simplicity complete markets and/or power utilities.

In the last chapter of this thesis we move on to the so-called *Principal-Agent Problem* (PA hereafter) of delegated portfolio management. As opposed to the previous sections, we restrict ourselves to discrete-time here. This problem is an instance of PA problems under moral hazard, and we refer to Holmström and Milgrom [1987], Schättler and Sung [1993] for some of the seminal contributions in this theory, Williams [2013] and the compendium Cvitanić and Zhang [2013] where mainly Pontryagin stochastic maximum principle is applied, Sannikov [2008] for a modern approach using as well Hamilton-Jacobi-Bellman (HJB) equations and discussing the relevant issues of retirement and quitting, among other types of contracting relationships, and Cvitanić et al. [2014] for a very recent approach to the case of moral hazard under incomplete information.

In our setting, however, we take as motivation the work of Ou-Yang [2003] (which in a sense is generalized in Cadenillas et al. [2007]), where a delegated portfolio management problem with linear contracts in continuous-time was analyzed, and look at its discrete-time variant as outlined in sections 4.2 and 4.4. Thought of as a PA problem under moral hazard, their setting is the following: an investor (the *Principal*) wants to get her capital invested by a manager (the *Agent*) in a financial market, for which a contract between them is to be designed so that it is in the latter's best interest to behave optimally for the former. The key issue is that the Principal cannot in principle force the Agent to choose what she wants and very often the latter's decisions are not observable nor contractible by the former. In the mentioned article and mainly under the assumption of exponential utilities the contracting problem is solved by means of a HJB approach and the optimal contract is showed to be of the form

"lump-sum payment plus gains/losses with respect to a benchmark portfolio,"

which is in fact the most common compensation structure used in practice. As it is the case for many specific PA models analysed in the literature, in Ou-Yang [2003] the author achieves to solve the problem explicitly to a significant extent and extract qualitative understanding of the situation. However when one considers a fully general PA problem (under moral hazard), things become much more entangled. For instance, in Cvitanić and Zhang [2013] the contracting problem reduces to solving a fully-coupled system of Forward-Backward Stochastic Differential Equations (FBSDE for short). Due to the generality of their setting, it is barely possible to gain an understanding out of such a system, and often it is not even well-known if the mentioned system admits a global solution at all. This trade-off between tractability and generality is a constant actor in the literature on PA problems. In this thesis we deal with this phenomenon by restricting ourselves to the discrete-time setting with linear contracts yet otherwise considering very arbitrary utility functions and price dynamics.

The main technical tool for our approach will be *Conditional Analysis*, which allows to translate most of the usual results in Analysis (therefore also convex analysis, optimization, etc.) to the case when sets and set-relations are replaced by suitable conditional variants thereof. We refer the reader to Filipović et al. [2012], Filipović et al. [2009] and Cheridito et al. [b] to get a flavour of the type of statements and results available in that theory, and to our section 4.5 for a summary of known results plus a few new ones which will be useful in our framework. The whole point is that after we write our dynamic optimization PA problem, and under suitable assumptions like *time-consistency* and *cash invariance*, we may reduce it by Theorem 4.4.1 to a series of static yet random optimization problems of risk-sharing type but under constraints. For such a deduction we employ the usual argument of turning the Agent's inter-temporal wealth as a driving variable, in spirit similar to Spear and Srivastava [1987]. We are hence lead to dealing with random optimization problems and here conditional analysis provides the framework and tools to tackle them in an elegant and general way, at the same time avoiding the usual measurable selection argument typically needed in such situations and which tend to be highly technical. We should stress that this programme has already been used in the different setting of equilibrium (see Cheridito et al. [a]).

The first main difficulty that arises in our approach is that at a first glance the mentioned random (i.e. conditional in our case) static optimization problems are not of a convex kind, owing to the Principal having to take account of the Agent's rationality into her decision making as a constraint; this is usually called incentive compatibility in the literature. However we very directly prove in Theorem 4.4.2 that both Principal's and Agent's problems can be merged into a single unconstrained one in our setting, whose solution yields the optimal contract. In economical terms, we see that an optimal first-best contract (i.e. one obtained as if the Principal could force the Agent to do what she wants) is implementable by the Agent, meaning that it is in his best interest to behave as the Principal wanted, and thus this first-best contract is also optimal in the original situation with moral hazard. This is already a generalization of the related result in Ou-Yang [2003] and is connected to Korn and Kraft [2008].

The second main difficulty is then solving the conditional optimization problems which together yield the first-best contract. These being unconstrained convex ones (strictly speaking concave, as we will be always maximizing), we are now in a better position to tackling them. The approach we follow is to prove that the set of potential optimizers is bounded in a suitable sense, and this is indeed obtained under several assumptions and in different contexts. In the greatest generality we work in the conditional version of L^1 spaces and with conditional utility functions enjoying a certain variational representation in the spirit of Maccheroni et al. [2006], and the transit from boundedness to optimality is achieved in Theorem 4.6.2 through either a *randomized Bolzano-Weierstrass Theorem* or a *Komlos-type argument* involving convex combinations of the original bounded, optimizing sequences. On the other extreme, and under suitable assumptions making it possible to essentially reduce the static conditional problems to deterministic ones in Euclidean spaces, we find in Theorem 4.8.1 the optimal contract by the Lagrange multiplier method. In the latter case we recover the known result that the optimal contract is of the form

> "lump-sum payment plus gains/losses with respect to a benchmark portfolio."

We close the chapter with a brief discussion on possible extensions and lines of research opened by the present work and a wrap-up discussion.

Acknowledgments

Hereby I would like to express my gratitude towards my supervisor Ulrich Horst for his guidance, support and continuous encouragement during my time as a PhD student, as well as to the Berlin Mathematical School for granting me full financial support and endless learning opportunities.

I also owe special thanks to Francisco Silva and Joaquín Fontbona for their friendship, support, enthusiasm and the joint work undertaken.

Regarding my colleagues, I am particularly thankful to Todor Bilarev, Jana Bielagk, Victor Fenou Nzengang, Paulwin Graewe, Asgar Jamnseshan, Oliver Janke, Martin Karliczek, Klebert Kentia Tonleu, Christoph Mainberger and Ludovic Tangpi, for many helpful discussions both during seminars and personal conversations.

My experience in Germany would not have been as enjoyable without the company and affection of my extended WG: Annika, Akiko, Caroline, Christian and Jochen.

Por último, y más importantemente, quiero agradecer el apoyo y amor incondicional de mi madre, sin cuyo ejemplo y dedicación yo no estaría donde estoy. También agradezco a mi polola Rayen por ser mi cable a tierra (ñuke mapu) y mejor regalo de año nuevo.

2 Robust utility maximization without model compactness

2.1 Introduction

In the robust utility maximization problem we deal with in this chapter, an investor/agent seeks to optimize her expected utility from final wealth taking into account the fact that she does not know accurately the "true" market dynamics. This is modelled as she taking a worst-case approach with respect to a family of possible probability measures which we call here the *uncertainty set*. As it was mentioned in Chapter 1, a prevalent assumption in the literature on the subject is to work with uncertainty sets with some sort of compactness property. In the present chapter we show how this assumption can be dispensed with, essentially by introducing a functional framework in a way that the images through the utility function of the terminal admissible wealths become (weakly) compact and thus allowing to relax the restrictions on the uncertainty set. Further, in the complete case, we provide a new characterization of the worst-case measures (i.e. those measures in the uncertainty set which yield the least optimal utility to the investor) by means of general convex duality. For pedagogical reasons we shall present our results in the complete and incomplete case rather separately, even though most of them in the former case are a direct consequence of those in the latter case, as notation and ideas are simpler to grasp in the complete case.

The chapter is organized as follows. In the next section, we describe the mathematical framework of the robust optimization problem in continuous-time financial markets and we recall the main results established in Schied and Wu [2005]. Then, we will state in a simplified way our main results about the incomplete market case. We further specialize our survey of results in the complete market case and illustrate their application with a simple example not covered by the previous literature. Then we end the section by working out such simple example, where our methodology provides in the complete case an explicit description of the worst-case measure and the optimal final wealth. In Section 2.3 we introduce and study some properties of the Orlicz-Musielak spaces that will be relevant in the complete case. Our main results on the robust optimization problem in that case (including the characterization of the worst-case measure) are then established in a general form in Section 2.4. In Section 2.5 we introduce the Modular spaces associated with the incomplete case. We then deduce from their study a new general minimax result which in particular entails the existence of optimal wealth processes. We also discuss the issue of reflexivity of our Modular spaces, proving that such a property seldom holds beyond the complete market case. We close the chapter with a rather heuristic discussion in section 2.6, where we address how our method could be extended to the case where there is no known global reference probability measure governing the market (in the sense that the uncertainty set is not absolutely continuous with respect to any measure a priori). Except for such section, this chapter is based on a joint work with Professor Joaquín Fontbona of the Universidad de Chile, which can be found in Backhoff and Fontbona and which is itself a deep extension of the present author's earlier work Backhoff.

2.2 Preliminaries and statement of main results

We will work in the same setting as Schied and Wu [2005], Kramkov and Schachermayer [1999]. Let there be d stocks and a bond, normalized to one for simplicity. Let $S = \left(S^i\right)_{1\leq i\leq d}$ be the price process of these stocks, and $T < \infty$ a finite investment horizon. The process S is assumed to be a semimartingale in a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F})_{t\leq T}, \mathbb{P})$, where $\mathbb{P}$ will always stand for the *reference measure*. The expectation with respect to $\mathbb{P}$ will be denoted by $\mathbb{E}$. The set of all probability measures on $(\Omega, \mathcal{F})$ absolutely continuous w.r.t. $\mathbb{P}$ will be denoted by $\mathcal{P}$, and the expectation with respect to $\mathbb{Q} \in \mathcal{P}\backslash\{\mathbb{P}\}$ will be expressed by $\mathbb{E}^{\mathbb{Q}}$.

A (self-financing) portfolio π is defined as a couple (X_0, H), where $X_0 \geq 0$ denotes the (constant) initial value associated to it and $H = (H^i)_{i=1}^d$ is a predictable and S-integrable process which represents the number of shares of each type under possession. The wealth associated to a portfolio π is the process $X = (X_t)_{t\leq T}$ given by

$$X_t = X_0 + \int_0^t H_u \mathrm{d}S_u, \tag{2.2.1}$$

and the set of attainable wealths from x is defined as

$$\mathcal{X}(x) = \{X \geq 0 : X \text{ as in (2.2.1) s.t. } X_0 \leq x\}. \tag{2.2.2}$$

The set of equivalent local martingale measures (or risk neutral measures) associated to S is

$$\mathcal{M}^e(S) = \{\mathbb{P}^* \sim \mathbb{P} : \text{ every } X \in \mathcal{X}(1) \text{ is a } \mathbb{P}^*\text{-local martingale}\} \tag{2.2.3}$$

which reduces to

$$\mathcal{M}^e(S) = \{\mathbb{P}^* \sim \mathbb{P} : S \text{ is a } \mathbb{P}^*\text{-local martingale}\},$$

if S is locally bounded. We assume this in the sequel, and that the market is *arbitrage-free* in the sense of NFLVR, meaning that $\mathcal{M}^e(S)$ is not empty.

As usual the market model is coined *complete* if $\mathcal{M}^e(S)$ is reduced to a singleton, i.e. $\mathcal{M}^e(S) = \{\mathbb{P}^*\}$. Given $\mathbb{Q} \in \mathcal{P}$, the following set generalizes the set of density processes (with respect to $\mathbb{Q}$) of risk neutral measures equivalent to it:

$$\mathcal{Y}_{\mathbb{Q}}(y) := \{Y \geq 0 | Y_0 = y \ , \ XY \text{ is } \mathbb{Q}-\text{supermartingale } \forall X \in \mathcal{X}(1)\} \, .$$

Introduced in Kramkov and Schachermayer [1999], $\mathcal{Y}_{\mathbb{Q}}(y)$ plays a central role in portfolio optimization in incomplete markets.

Definition 2.2.1 *A function $U : (0, \infty) \to \mathbb{R}$ is called a utility function on $(0, +\infty)$, if it is strictly increasing, strictly concave and continuously differentiable. It will be said to satisfy INADA if*

$$U'(0+) = \infty \text{ and } U'(+\infty) = 0 \, .$$

Its asymptotic elasticity, introduced in Kramkov and Schachermayer [1999], is defined as $AE(U) := \limsup_{x\to\infty} \frac{xU'(x)}{U(x)}$.

Such a function U is extended as $-\infty$ on $(-\infty, 0)$.

Suppose now that an agent aims to optimize her wealth by investing in a market which might be modelled by more than one probabilistic model, the actual or more accurate one being unknown to her. Let $\mathcal{Q} \subset \mathcal{P}$ be a set of feasible probability measures or models on $(\Omega, \mathcal{F}, (\mathcal{F})_{t\leq T}, \mathbb{P})$ representing the mentioned ambiguity or uncertainty. We shall refer to such a set as the *uncertainty set* from here on. A common paradigm in robust optimization consists in adopting a conservative or risk averse point of view, in which the agent tries to solve the optimization problem

$$\sup_{X\in\mathcal{X}(x)} \inf_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}} \left(U \left(X_T \right) \right), \tag{2.2.4}$$

(a suitable meaning can often be given to the expectation in case U is unbounded) which represents the situation in which she tries to maximize the worst-case expected utility given the set of models under consideration.

Throughout the present work it will be assumed that $\mathcal{Q}$ contains only probability measures that are absolutely continuous with respect to $\mathbb{P}$. We will write

$$\mathcal{Q}_e := \{\mathbb{Q} \in \mathcal{Q} | \mathbb{Q} \sim \mathbb{P}\},$$

and respectively denote by $\frac{d\mathcal{Q}}{d\mathbb{P}}$ and $\frac{d\mathcal{Q}_e}{d\mathbb{P}} \subset \frac{d\mathcal{Q}}{d\mathbb{P}}$ the set of densities with respect to $\mathbb{P}$ of the elements of $\mathcal{Q}$ and $\mathcal{Q}_e$:

$$\frac{d\mathcal{Q}}{d\mathbb{P}} := \left\{ \frac{d\mathbb{Q}}{d\mathbb{P}} : \mathbb{Q} \in \mathcal{Q} \right\}, \quad \frac{d\mathcal{Q}_e}{d\mathbb{P}} := \left\{ \frac{d\mathbb{Q}}{d\mathbb{P}} : \mathbb{Q} \in \mathcal{Q}_e \right\} = \left\{ \frac{d\mathbb{Q}}{d\mathbb{P}} \in \frac{d\mathcal{Q}}{d\mathbb{P}} : \frac{d\mathbb{Q}}{d\mathbb{P}} > 0 \text{ a.s. } \right\}.$$

As in the standard, i.e. non-robust, setting (see Pham [2009] for general background) the dual formulation of the optimization problem (2.2.4) will make use of the conjugate function of U (actually the Fenchel conjugate of $-U(-\cdot)$), given by

$$V(y) := \sup_{x>0} [U(x) - xy] \ \forall y > 0. \tag{2.2.5}$$

The following functions commonly used in the literature to tackle problem (2.2.4), will also be relevant here:

$$
\begin{aligned}
u(x) &= \sup_{X \in \mathcal{X}(x)} \inf_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}^{\mathbb{Q}} \left(U \left(X_T \right) \right), && (2.2.6) \\
u_{\mathbb{Q}}(x) &= \sup_{X \in \mathcal{X}(x)} \mathbb{E}^{\mathbb{Q}} \left(U \left(X_T \right) \right), && (2.2.7) \\
v_{\mathbb{Q}}(y) &= \inf_{Y \in \mathcal{Y}_{\mathbb{Q}}(y)} \mathbb{E}^{\mathbb{Q}} \left(V \left(Y_T \right) \right), && (2.2.8) \\
v(y) &= \inf_{\mathbb{Q} \in \mathcal{Q}_e} v_{\mathbb{Q}}(y). && (2.2.9)
\end{aligned}
$$

Of course, $u_{\mathbb{Q}}(x)$ is the investor's "subjective" utility under model $\mathbb{Q} \in \mathcal{Q}_e$, when starting from an initial wealth not larger that $x > 0$, whereas $u(x)$ is her robust utility. The function $x \mapsto u_{\mathbb{Q}}(x)$ is concave (as an easy check shows), so that $u_{\mathbb{Q}}(x_0) < +\infty$ at some $x_0 > 0$ for some given $\mathbb{Q} \in \mathcal{Q}$ implies $u_{\mathbb{Q}} < +\infty$ and then, $u < +\infty$, by the usual minimax inequality.

For a fixed $\mathbb{Q} \in \mathcal{Q}_e$ it was proven in [Kramkov and Schachermayer, 1999, Theorem 3.1] that, whenever $u_{\mathbb{Q}}$ is finite, the functions $u_{\mathbb{Q}}$ and $v_{\mathbb{Q}}$ are conjugate:

$$
u_{\mathbb{Q}}(x) = \inf_{y>0} \left(v_{\mathbb{Q}}(y) + xy \right) \text{ and } v_{\mathbb{Q}}(y) = \sup_{x>0} \left(u_{\mathbb{Q}}(x) - xy \right). \tag{2.2.10}
$$

Hence, since the inequalities

$$
\begin{aligned}
u(x) &\leq \inf_{y>0} \left(\inf_{\mathbb{Q} \in \mathcal{Q}} \inf_{Y \in \mathcal{Y}_{\mathbb{Q}}(y)} \mathbb{E}^{\mathbb{Q}} \left(V \left(Y_T \right) \right) + xy \right) \\
&\leq \inf_{y>0} \left(\inf_{\mathbb{Q} \in \mathcal{Q}_e} \inf_{Y \in \mathcal{Y}_{\mathbb{Q}}(y)} \mathbb{E}^{\mathbb{Q}} \left(V \left(Y_T \right) \right) + xy \right) = \inf_{y>0} \left(v(y) + xy \right),
\end{aligned} \tag{2.2.11}
$$

always hold, the function v can be considered as a candidate conjugate of u.

We will denote in the sequel by $L^0 = L^0(\Omega, \mathbb{P})$ the space of measurable functions equipped with the topology of convergence in probability, and by $L^0_+ \subset L^0$ the cone of non-negative functions therein.

Let us now briefly summarize the main available general results on the robust problem, obtained in Schied and Wu [2005]. The following assumption on $\mathcal{Q}$ is required:

Assumption 1

1. *$\mathcal{Q}$ is convex.*

2. *$\mathbb{P}(A) = 0$ if and only if $[\mathbb{Q}(A) = 0, \forall \mathbb{Q} \in \mathcal{Q}]$.*

3. *The set $\frac{\mathrm{d}\mathcal{Q}}{\mathrm{d}\mathbb{P}}$ is closed in $L^0(\mathbb{P})$, i.e. with respect to convergence in $\mathbb{P}$-measure.*

Theorem 2.2.1 (Theorem 2.2,Schied and Wu [2005]) *Suppose Assumptions 1 and $\mathcal{M}^e(S) \neq \emptyset$ hold, as well as:*

$$
\exists x > 0, \mathbb{Q}_0 \in \mathcal{Q}_e \ st. \ u_{\mathbb{Q}_0}(x) < \infty. \tag{2.2.12}
$$

Then the function u defined in (2.2.6) *is concave, finite, and satisfies the minimax identity:*

$$u(x) = \inf_{\mathbb{Q}\in\mathcal{Q}} \sup_{X\in\mathcal{X}(x)} \mathbb{E}^{\mathbb{Q}}\left(U\left(X_T\right)\right). \tag{2.2.13}$$

Moreover, u and v as in (2.2.9) *are conjugate:*

$$u(x) = \inf_{y>0}\left(v(y) + xy\right) \ , \text{ and } v(y) = \sup_{x>0}\left(u(x) - xy\right). \tag{2.2.14}$$

In particular, v is convex. Also, their derivatives satisfy:

$$u'(0+) = \infty \ , \text{ and } , \ v'(\infty-) = 0. \tag{2.2.15}$$

Theorem 2.2.2 (Theorem 2.6,Schied and Wu [2005]) *Suppose Assumption 1 and*

$$\forall y > 0, \forall \mathbb{Q} \in \mathcal{Q}_e, \quad v_{\mathbb{Q}}(y) < \infty \tag{2.2.16}$$

(which is true as soon as $u_{\mathbb{Q}}$ is finite $\forall \mathbb{Q} \in \mathcal{Q}_e$ and $AE(U) < 1$). Then, the derivatives of the value functions satisfy:

$$v'(0+) = -\infty \ , \text{ and } u'(\infty-) = 0, \tag{2.2.17}$$

and $\forall x > 0$, $\exists \hat{X} \in \mathcal{X}(x)$ and a measure $\hat{\mathbb{Q}} \in \mathcal{Q}$ such that:

$$u(x) = \inf_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}}\left[U\left(\hat{X}_T\right)\right] = \mathbb{E}^{\hat{\mathbb{Q}}}\left[U\left(\hat{X}_T\right)\right] = u_{\hat{\mathbb{Q}}}(x), \tag{2.2.18}$$

that is, the suprema and infima in (2.2.13) *are attained. Moreover, there exists $\hat{y}$ in the superdifferential of u at x, and some $\hat{Y} \in \mathcal{Y}_{\mathbb{P}}(\hat{y})$ such that:*

$$v(\hat{y}) = \mathbb{E}\left[\hat{Z} V\left(\frac{\hat{Y}_T}{\hat{Z}}\right)\right] \ , \text{ and } , \ \hat{X}_T = [U']^{-1}\left(\frac{\hat{Y}_T}{\hat{Z}}\right) \ \hat{\mathbb{Q}} - a.e., \tag{2.2.19}$$

where $\hat{Z} = \frac{\mathrm{d}\hat{\mathbb{Q}}}{\mathrm{d}\mathbb{P}}$. What is more, $\hat{X}\hat{Y}$ is a $\mathbb{P}$-martingale and v satisfies:

$$v(y) = \inf_{\mathbb{M}\in\mathcal{M}^e(S)} \inf_{\mathbb{Q}\in\mathcal{Q}_e} \mathbb{E}^{\mathbb{Q}}\left[V\left(y\frac{\mathrm{d}\mathbb{M}}{\mathrm{d}\mathbb{Q}}\right)\right]. \tag{2.2.20}$$

If additionally $AE(U) < 1$, then u is strictly concave, v is continuously differentiable, and:

$$\left\{\hat{X}_T\hat{Y}_T > 0\right\} = \left\{\hat{Z} > 0\right\} \ \mathbb{P} - a.e. \tag{2.2.21}$$

Some comments about Assumption 1 on $\mathcal{Q}$ are in order. Point (1) together with (3) imply that $\mathcal{Q}$ is countably convex, and together with point (2) this is used in Schied and Wu [2005] to ensure that $\mathcal{Q}_e$ is not empty (thanks to Halmos-Savage Theorem, see [Klein and Schachermayer, 1996, Theorem 1.1] or [Föllmer and Schied, 2004, Theorem 1.61]).

More importantly, in view of points (1) and (2), point (3) is equivalent to $\frac{\mathrm{d}\mathcal{Q}}{\mathrm{d}\mathbb{P}}$ being a $\sigma(L^1, L^\infty)$-*compact set* (see Lemma 3.2 of Schied and Wu [2005]). This fact turns out to be crucial in the proofs of the above results, in order to establish, among other things, the minimax identity (2.2.13) and the expression for $v(\hat{y})$ in (2.2.19), as well as to ensure the fact that the double infimum in (2.2.9) is attained. To our knowledge, the same L^1- weak compactness condition is present, for instance in Föllmer and Gundel [2006], where the authors study the above problem through a different approach (of robust projections) and, in some way or another, in all the available results about problem (2.2.4).

The next example shows, however, that meaningful uncertainty sets which are not closed in L^0 arise naturally or are simple to conceive:

Example 2.2.1 *Let us imagine that the investor in the continuous time market model* (2.2.1) *has an apriori knowledge (as in insider trading) or belief (as in our robustness interpretation) that on average a certain $\mathcal{F}_T-$measurable unbounded random variable h (e.g. S_T) is bounded from below by a constant $A > 0$. If $\mathbb{E}(h) < \infty$, then the set of densities $\frac{\mathrm{d}\mathcal{Q}_A}{\mathrm{d}\mathbb{P}}$ of the uncertainty set $\mathcal{Q}_A := \{\mathbb{Q} \in \mathcal{P} : \mathbb{Q} \ll \mathbb{P}, \mathbb{E}^{\mathbb{Q}}(h) \geq A\}$ is not-closed in L^0. Indeed, the sequence $\mathbb{Q}^n(\cdot) := \mathbb{P}(\cdot|h \geq nA) \in \mathcal{Q}_A$, is such that $\frac{\mathrm{d}\mathbb{Q}^n}{\mathrm{d}\mathbb{P}} = \mathbb{P}(h \geq nA)^{-1}\mathbb{1}_{\{h\geq nA\}} \to 0$ in L^0 when $n \to \infty$, yet obviously $0 \notin \mathcal{Q}_A$.*

Our main goal is to establish a functional framework allowing us to study the robust optimization problem without the L^1- weak compactness assumption, and to recover at least in some general situations, some of the results in Schied and Wu [2005] in such a setting. The spaces and tools we will introduce will be naturally related to the elements of the problem, and they will allow us to deal with some examples of uncertainty sets $\mathcal{Q}$ that commonly arise in concrete situations.

In the remainder of this chapter, we will restrict our attention to the setting of

Assumption 2 *U is a utility function on $(0, \infty)$, not bounded from above, satisfying INADA and such that $U(0+) = 0$.*

Remark 2.2.1 *It is easy to see that power utilities (i.e. $U(\cdot) = \alpha^{-1}(\cdot)^\alpha$, $\alpha \in (0, 1)$) fulfil such set of conditions. Moreover, the above assumption is satisfied if and only if for the inverse of U it holds: U^{-1} is convex and increasing, $U^{-1}(0+) = 0$, $U^{-1}(\infty) = \infty$, $[U^{-1}]'(0+) = 0$ and $[U^{-1}]'(\infty) = \infty$. With this we can see that, for instance, the inverse on $[0, +\infty)$ of $x \mapsto e^x - x - 1$ satisfies Assumption 2.*

Remark 2.2.2 *If $U(0+) > -\infty$ only, by a translation argument it can be assumed w.l.o.g. that $U(0+) = 0$. Also, under the latter condition we have $V \geq 0$.*

An overview of our approach and results is presented in the following subsections.

2.2.1 Main statements in general markets

Suppose for ease of exposition that the reference measure is a martingale one. The candidate conjugate to u is:

$$v(y) = \inf_{Z \in \frac{\mathrm{d}\mathcal{Q}_e}{\mathrm{d}\mathbb{P}}} \inf_{Y \in \mathcal{Y}} \mathbb{E}\left[ZV\left(\frac{yY}{Z}\right)\right], \tag{2.2.22}$$

where $\mathcal{Y} := \mathcal{Y}_{\mathbb{P}}(1)$, and we often write Y for Y_T. Therefore, if equality at some finite value is to hold in (2.2.11), the optimization problem (2.2.4) can be restricted to measures $\mathbb{Q} \in \mathcal{Q}$ for which $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}$ is in the space of measurable functions

$$L_I := \bigcup_{Y \in \mathcal{Y}} L_{|\cdot| V \circ Y/|\cdot|},$$

where for every $Y \in \mathcal{Y}$ we define:

$$L_{|\cdot| V \circ Y/|\cdot|} := \left\{ Z \in L^0 \text{ s.t. } \exists \alpha > 0, \mathbb{E}^{\mathbb{P}}\left[|Z| V(Y/(\alpha|Z|))\right] < \infty \right\}.$$

The function $z \mapsto |z| V(Y/|z|)$ is a.s. non-negative and convex under Assumption 2, so that $L_{|\cdot| V \circ Y/|\cdot|}$ will turn out to be an Orlicz-Musielak space (see Remark 2.3.3), hence a Banach space with the adequate norms. The convex conjugate of $|\cdot| V \circ Y/|\cdot|$ will be shown to be the function $YU^{-1} \circ |\cdot|$, and it will play a pre-eminent role, as will do the associated Orlicz-Musielak space

$$L_{YU^{-1} \circ |\cdot|} := \left\{ Z \in L^0 \text{ s.t. } \exists \alpha > 0, \mathbb{E}\left[YU^{-1}(\alpha|Z|)\right] < \infty \right\}.$$

Relevant properties of $L_{|\cdot| V \circ Y/|\cdot|}$ and $L_{YU^{-1} \circ |\cdot|}$ will be pointed out in a more general setting below. In particular the following conditions will be relevant in the study of topological duality between these spaces:

Assumption 3 *Assumption 2 on the utility function U holds and, for some constants $a, b, k, d > 0$, the convex functions $V(y) = \sup_{x>0}[U(x) - xy]$ and $U^{-1}(y)$ on $(0, \infty)$ satisfy*

$$V(y/2) \leq aV(y) + b(y+1) \;\forall y > 0, \tag{2.2.23}$$

and

$$U^{-1}(2y) \leq kU^{-1}(y) + d \;\forall y > 0. \tag{2.2.24}$$

In the jargon of Orlicz space theory (see e.g. Rao and Ren [1991]), Assumption 3 will correspond to "Δ_2 and ∇_2"-type conditions on the Young function $|\cdot| V \circ 1/|\cdot|$. Let us point out that this is satisfied for instance by the utility functions on $(0, \infty)$ given by $U(x) = \frac{x^\alpha}{\alpha}$, $\alpha \in (0, 1)$.

In Section 2.5.1 a suitable Banach Space topology on L_I is defined (called a Modular Space topology), which is a generalization of the Orlicz-Musielak one. Furthermore, we shall find that this norm topology harmonizes tightly with our optimization problems. We are thus led to finding verifiable conditions on the utility function U that may render the space L_I to be tractable. This is done for the next result, where under the right assumptions that allow us to identify the dual of L_I (with some concrete space L_J related to the intersection of the $L_{YU^{-1}\circ|\cdot|}$ spaces), we can obtain the minimax equality and existence of optimal strategies by exploiting a certain compactness of the image under U of the terminal wealths as elements in the dual space of L_I. This is the content of Theorem 2.5.1, of which we give a simplified version now:

Theorem 2.2.3 *Suppose Assumption 2, that (for simplicity) the reference measure* $\mathbb{P}$ *is already a martingale one, and that the set* $\mathcal{Q}$ *satisfies:*

- $\mathcal{Q}$ *is countably convex,*
- $[\mathbb{P}(A) = 0 \iff \forall \mathbb{Q} \in \mathcal{Q}, \mathbb{Q}(A) = 0]$,
- $\frac{\mathrm{d}\mathcal{Q}}{\mathrm{d}\mathbb{P}} \cap L_I$ *is non-empty and closed w.r.t. the topology on* L_I *weakened by its dual,*
- $\exists x_0 > 0, \exists \mathbb{Q}_0 \in \mathcal{Q}_e$ *such that* $u_{\mathbb{Q}_0}(x_0) < \infty$.

Then under condition (2.2.23) *in Assumption 3, we have that for every* $x > 0$*:*

$$\begin{aligned} u(x) \quad &= \inf_{\mathbb{Q}\in\mathcal{Q}} \sup_{X\in\mathcal{X}(x)} \mathbb{E}^{\mathbb{Q}}\left(U\left(X_T\right)\right) \quad = \inf_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}}\left(U\left(\hat{X}_T\right)\right) \\ &= \inf_{\mathbb{Q}\in\mathcal{Q}_e} \sup_{X\in\mathcal{X}(x)} \mathbb{E}^{\mathbb{Q}}\left(U\left(X_T\right)\right) \quad < +\infty, \end{aligned} \tag{2.2.25}$$

for some $\hat{X} \in \mathcal{X}(x)$*. Moreover* v *is finite and* u, v *are conjugates on* $(0, \infty)$*.*

In Section 2.5.2 we will build up the rigorous functional analytic setting in order to prove the above result. In the section thereafter we will further see that a sufficient condition for the existence of a saddle point (hence a worst-case measure) is that L_I be a reflexive space, which is why we also investigate conditions for that property to hold. The main result in this respect, stated next, gives a rather sobering answer to that question:

Theorem 2.2.4 *Under Assumptions 2 and 9, if the set* $\mathcal{Y}$ *is not uniformly integrable, then* L_I *is not reflexive.*

As it shall be discussed, in most reasonable strict incomplete market models (for instance those involving the brownian filtration) the mentioned set is not uniformly integrable and thus L_I is not reflexive. On the positive side, in the complete case $\mathcal{Y}$ is of course dominated in L^1 (see e.g. [Kramkov and Schachermayer, 1999, Lemma 4.3]) and therefore the previous result does not exclude reflexivity in that case. We will actually see that under Assumption 3 the space L_I is reflexive in the complete market case. This fact will allow us to fully remove in the complete case the assumption of L^1-weakly compact uncertainty sets, recover in that enlarged setting the main statement of Theorems 2.2.1 and 2.2.2, and state new results characterizing the worst-case measure.

2.2.2 Main statements in the complete market case

We specify the discussion to the complete setting, otherwise keeping the notation introduced so far. In this case, we shall have

$$v(y) = \inf_{Z \in \frac{\mathrm{d}\mathcal{Q}_e}{\mathrm{d}\mathbb{P}}} \mathbb{E}\left[ZV\left(\frac{y}{Z}\right)\right]. \tag{2.2.26}$$

Then, if equality at some finite value is to hold in (2.2.11), the optimization problem (2.2.4) can be restricted to measures $\mathbb{Q} \in \mathcal{Q}$ for which $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}$ is in the space of measurable functions

$$L_{|\cdot|V \circ 1/|\cdot|} := \left\{Z \in L^0 \text{ s.t. } \exists \alpha > 0, \mathbb{E}\left[|Z| V(1/(\alpha |Z|))\right] < \infty\right\}.$$

The space $L_{|\cdot|V \circ 1/|\cdot|}$ is a classical Orlicz space, and in the current setting it coincides, as a topological space, with the space L_I previously introduced.

Because in the complete case we can sharpen our results, in particular providing existence and characterization of worst-case measures and optimal strategies, we shall write in detail the assumptions and results that we need and obtain:

Assumption 4

- *$\mathcal{Q}$ is countably convex.*
- $[\mathbb{P}(A) = 0 \iff \forall \mathbb{Q} \in \mathcal{Q}, \mathbb{Q}(A) = 0]$.
- $\frac{\mathrm{d}\mathcal{Q}}{\mathrm{d}\mathbb{P}} \cap L_{|\cdot|V \circ 1/|\cdot|}$ *is a non-empty, weakly closed convex set of* $L_{|\cdot|V \circ 1/|\cdot|}$
- $\exists x_0 > 0, \exists \mathbb{Q}_0 \in \mathcal{Q}_e$ *such that* $u_{\mathbb{Q}_0}(x_0) < \infty$,

As in the assumptions in Theorem 2.2.3, and unlike Assumption 1, the third condition depends on the utility function at hand. Since we cannot get countable convexity out of convexity in the present context, we add this to the assumptions. The fourth condition, which we add straight from the beginning, is required in any case for most of the results in Schied and Wu [2005].

We state now our main result in the complete case, which will be proved in Section 2.4.1. We phrase it on purpose as in the corresponding results in Schied and Wu [2005]:

Theorem 2.2.5 *Assume that the market is complete, and (only for simplicity) that the reference measure $\mathbb{P}$ is the risk-neutral one. Suppose Assumptions 3 and 4 hold. Then:*

a) The function u is concave, finite, and satisfies the minimax identity

$$u(x) = \inf_{\mathbb{Q} \in \mathcal{Q}} \sup_{X \in \mathcal{X}(x)} \mathbb{E}^{\mathbb{Q}}\left(U\left(X_T\right)\right).$$

What is more, v is finite, convex l.s.c, and u, v are conjugates on $(0, \infty)$*:*

$$u(x) = \inf_{y>0} \left(v(y) + xy\right) \ , \ v(y) = \sup_{x>0} \left(u(x) - xy\right).$$

b) For every $x > 0$ there exists a measure $\hat{\mathbb{Q}} \in \mathcal{Q}$ and a $\mathbb{P}-$martingale $\hat{X} \in \mathcal{X}(x)$ such that:

$$u(x) = \inf_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}^{\mathbb{Q}} \left[U \left(\hat{X}_T \right) \right] = \mathbb{E}^{\hat{Q}} \left[U \left(\hat{X}_T \right) \right] = u_{\hat{\mathbb{Q}}}(x) = v(\hat{y}) + x\hat{y}, \qquad (2.2.27)$$

where $\hat{y}$ belongs to the super-differential of u at x, and

$$v(\hat{y}) = \mathbb{E}^{\hat{\mathbb{Q}}} \left[V \left(\hat{y} \left[\frac{\mathrm{d}\hat{\mathbb{Q}}}{\mathrm{d}\mathbb{P}} \right]^{-1} \right) \right],$$

as well as $\hat{\mathbb{Q}}-a.s.$

$$\hat{X}_T = [U']^{-1} \left(\hat{y} \left[\frac{\mathrm{d}\hat{\mathbb{Q}}}{\mathrm{d}\mathbb{P}} \right]^{-1} \right).$$

c) If additionally $AE(U) < 1$, then u is strictly concave, v is continuously differentiable, and $\mathbb{P}-a.s.$ one has $\hat{X}_T = [U']^{-1} \left(\hat{y} \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\hat{\mathbb{Q}}} \right)$.

Remark 2.2.3 *By Hahn-Banach theorem, $\mathcal{Q} \subset \mathcal{P}$ satisfies the third point of Assumption 4 if and only if there exists a family $\mathcal{H} = (h_\lambda)_{\lambda \in \Lambda}$ of elements of $L_{U^{-1} \circ |\cdot|}$ and a function $\lambda \mapsto (a_\lambda, b_\lambda)$ with $-\infty \le a_\lambda \le b_\lambda \le +\infty$ such that*

$$\frac{\mathrm{d}\mathcal{Q}}{\mathrm{d}\mathbb{P}} \cap L_{|\cdot| V \circ 1/|\cdot|} = \bigcap_{\lambda \in \Lambda} \left\{ \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} \ : \ \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} \in L_{|\cdot| V \circ 1/|\cdot|} \text{ and } \mathbb{E}^{\mathbb{Q}}(h_\lambda) \in [a_\lambda, b_\lambda] \right\}. \qquad (2.2.28)$$

The functions h_λ in Remark 2.2.3 can actually be interpreted as real “observables of the market,” so that the uncertainty set can always be understood as those models under which their expected observed values $\mathbb{E}^{\mathbb{Q}}(h_\lambda)$ lie, when defined, on the prescribed extended real intervals $[a_\lambda, b_\lambda]$. Uncertainty sets specified in such way naturally arise in modelling situations (e.g. information on moments).

We will later see that under Assumption 2 on the utility function, the first three points in Assumption 4 on the uncertainty set are implied by Assumption 1. It is easy to see that the converse is not true, as we show in this example:

Example 2.2.2 *Consider the utility function $U(x) = \frac{x^\alpha}{\alpha}$, $\alpha \in (0, 1)$, so that $L_{|\cdot| V \circ 1/|\cdot|} = L^{\frac{1}{\alpha}}$, and the uncertainty set $\mathcal{Q}_A$ of Example 2.2.1 . If the r.v. h is in $L^{\frac{1}{1-\alpha}}$, one can check with help of Hölder's inequality that $L_{|\cdot| V \circ 1/|\cdot|} \cap \mathcal{Q}_A$ is a closed subset of $L_{|\cdot| V \circ 1/|\cdot|}$.*

2.2.3 Characterization of the solution in the complete case

Our next aim is to characterize the solution $\hat{\mathbb{Q}}$ of the robust portfolio optimization problem (i.e. the worst-case measure) in the complete case, by adapting to the present framework techniques developed in the context of abstract entropy minimization problems in a series of papers by C. Léonard (see Léonard [2008], Léonard [2010], Léonard [2003] and references therein). We will state in a particular (simplified) setting our main result on the characterization of the worst-case measure in the complete case. Some additional notation and hypotheses are needed (see Remark 2.2.3 for the context). We denote by $\mathcal{C}^\Lambda$ the convex subset of $\mathbb{R}^\Lambda$

$$\mathcal{C}^\Lambda := \{t \in \mathbb{R}^\Lambda : \forall \lambda \in \Lambda,\ t_\lambda \in [a_\lambda, b_\lambda]\}.$$

The following condition of linear independence regarding the family of observables $\mathcal{H}$ (enlarged with the constant observable 1) will be useful.

Assumption 5 *There exists a family of random variables $\mathcal{H} = (h_\lambda)_{\lambda\in\Lambda}$ associated with $\mathcal{Q}$ as in* (2.2.28)*, such that for each finite subset $\Lambda' \subset \Lambda$, and every $\alpha = (\alpha_\lambda) \in \mathbb{R}^{\Lambda'}$ and $\beta \in \mathbb{R}$,*

$$\beta + \sum_{\lambda\in\Lambda'} \alpha_\lambda h_\lambda = 0 \quad \mathbb{P} - \textit{a.s. if and only if } \alpha = 0 \textit{ and } \beta = 0.$$

It will be seen later on that Assumption 5 is not an actual restriction.

For each $y > 0$, we next introduce the function $\nu_y : \mathbb{R}^\Lambda \times \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ defined at $(t, s) = ((t_\lambda)_{\lambda\in\Lambda}, s)$ by

$$\nu_y(t,s) := \sup_{\Lambda'\subset\Lambda, |\Lambda'|<\infty} \sup_{\beta\in\mathbb{R}} \sup_{\alpha\in\mathbb{R}^{\Lambda'}} \beta s + \sum_{\lambda\in\Lambda'} \alpha_\lambda t_\lambda - y\mathbb{E}\left[U^{-1}\left(\left(\beta + \sum_{\lambda\in\Lambda'} \alpha_\lambda h_\lambda\right)_+\right)\right].$$

The following is a key assumption (introduced in Léonard [2008]) to be interpreted as a qualification condition of weak type which enables the characterization of the minimizing measures. Recall that the *affine hull* aff(A) of $A \subset L$, where L is a linear space, is the smallest affine subspace of L containing A, and the *intrinsic core* of A is

$$\mathrm{icor}(A) := \{a \in A | \forall x \in \mathrm{aff}(A), \exists t > 0 \text{ st. } a + t(x - a) \in A\};$$

this is the biggest topology-free definition of the interior of a set.

Assumption 6 *For each $y > 0$:*

$$\left(\mathcal{C}^\Lambda \times \{1\}\right) \cap \ \textit{icor}\ (\textit{dom}\,\nu_y) \neq \emptyset.$$

We also write $\mathbb{R}^\Lambda_\mathcal{H}$ for the linear subspace of $\mathbb{R}^\Lambda \times \mathbb{R}$ given by

$$\mathbb{R}^\Lambda_\mathcal{H} := \{(t,s) \in \mathbb{R}^\Lambda \times \mathbb{R} : \exists Z \in L_{|\cdot|V\circ 1/|\cdot|} \text{ s.t. } t_\lambda = \mathbb{E}(Zh_\lambda)\ \forall\lambda \in \Lambda, s = \mathbb{E}(Z)\},$$

the expectations making sense by Hölder's inequality in Orlicz spaces. Notice that for each $(t,s) = (\mathbb{E}(Zh_\lambda)_{\lambda\in\Lambda}, \mathbb{E}(Z)) \in \mathbb{R}^\Lambda_\mathcal{H}$, the linear mapping defined on the span $\mathbb{R}+\langle\mathcal{H}\rangle = \{\beta + \sum_{\lambda\in\Lambda'}\alpha_\lambda h_\lambda : \beta\in\mathbb{R}, (\alpha_\lambda)\in\mathbb{R}^{\Lambda'}, \Lambda'\subset\Lambda, |\Lambda'|<\infty\}$ by

$$\beta + \sum_{\lambda\in\Lambda'}\alpha_\lambda h_\lambda \mapsto \beta s + \sum_{\lambda\in\Lambda'}\alpha_\lambda t_\lambda = \mathbb{E}\left(Z\left(\beta + \sum_{\lambda\in\Lambda'}\alpha_\lambda h_\lambda\right)\right),$$

can be extended by density to a unique linear map defined in the closure $\overline{\mathbb{R}+\langle\mathcal{H}\rangle}$ of $\mathbb{R}+\langle\mathcal{H}\rangle$ in $L_{U^{-1}\circ|\cdot|}$ and denoted

$$\beta + h \mapsto \beta s + \langle h, t\rangle.$$

The definition does not depend on Z such that $(t,s) = (\mathbb{E}(Zh_\lambda)_{\lambda\in\Lambda}, \mathbb{E}(Z))$. We have

Theorem 2.2.6 *Suppose that the general assumptions of Theorem 2.2.5 hold, together with Assumption 5.*

a) For each $y>0$, the following identities hold:

$$\begin{aligned}
v(y) &= \inf_{t\in\mathcal{C}^\Lambda}\nu_y(t,1) = \inf_{t:(t,1)\in(\mathcal{C}^\Lambda\times\{1\})\cap\mathbb{R}^\Lambda_\mathcal{H}}\nu_y(t,1)\\
&= \sup_{\substack{\Lambda'\subset\Lambda,|\Lambda'|<\infty\\ \beta\in\mathbb{R},\alpha\in\mathbb{R}^{\Lambda'}}}\left(\inf_{t\in\mathcal{C}^\Lambda}\beta + \sum_{\lambda\in\Lambda'}\alpha_\lambda t_\lambda\right) - y\mathbb{E}\left[U^{-1}\left(\left(\beta + \sum_{\lambda\in\Lambda'}\alpha_\lambda h_\lambda\right)_+\right)\right]\\
&= \sup_{\beta+h\in\overline{\mathbb{R}+\langle\mathcal{H}\rangle}}\left(\inf_{t:(t,1)\in(\mathcal{C}^\Lambda\times\{1\})\cap\mathbb{R}^\Lambda_\mathcal{H}}\beta + \langle h,t\rangle\right) - y\mathbb{E}\left[U^{-1}\left((\beta+h)_+\right)\right].
\end{aligned} \tag{2.2.29}$$

Moreover, the infimum (2.2.26) *is attained at a unique element $Z^y \in \frac{\mathrm{d}\mathcal{Q}}{\mathrm{d}\mathbb{P}}$.*

b) For all $x>0$, we have:

$$u(x) = \inf_{y>0}\left(\inf_{t\in\mathcal{C}^\Lambda}\nu_y(t,1) + xy\right) = \inf_{y>0}\left(\mathbb{E}\left[Z^yV\left(\frac{y}{Z^y}\right)\right] + xy\right) = \mathbb{E}\left[Z^{\hat y}V\left(\frac{\hat y}{Z^{\hat y}}\right)\right] + x\hat y,$$

where $\hat y$ belongs to the super-differential of u at x.

c) If in addition Assumption 6 holds, then the second maximization problem in (2.2.29) *has a solution $\beta + h \in \overline{\mathbb{R}+\langle\mathcal{H}\rangle}$. Moreover, there exists a $\mathbb{P}-$a.s. unique non-negative function $\bar h \in L_{U^{-1}\circ|\cdot|}$ such that $(\beta+h)_+ = \bar h$ for any solution $\beta + h \in \overline{\mathbb{R}+\langle\mathcal{H}\rangle}$ of* (2.2.29)*, and the unique solution $Z^y\in\frac{\mathrm{d}\mathcal{Q}}{\mathrm{d}\mathbb{P}}$ of problem* (2.2.26) *is given by*

$$Z^y := y(U^{-1})'\big((\beta+h)_+\big) = y(U^{-1})'\big(\bar h\big).$$

Plainly, the previous result states that under suitable conditions, the problem of finding the worst-case measure in the robust portfolio optimization problem can (at least theoretically) be solved in the complete case, through the following strategy:

- finding for each $y > 0$ a solution $\beta + h$ to the last problem in (2.2.29)
- computing for such y the value $v(y) = \mathbb{E}\left[Z^y V\left(\frac{y}{Z^y}\right)\right]$, where

$$Z^y := y(U^{-1})'\big((\beta + h)_+\big),$$

- and minimizing on $y > 0$ the obtained values of $v(y) + xy$. Then, $Z^{\hat{y}}$ associated with the minimizer $\hat{y}$ is the worst-case measure.

Notice that, in general, $Z^{\hat{y}}$ might depend on x and on the utility function, contrary to the least favourable measures determined for instance in Baudoin [2003] or Schied [2005] (see also Föllmer and Schied [2004]) for specific uncertainty sets.

Of course, for each $y > 0$ the problem (2.2.29) is a dual problem to (2.2.26) and so, in some sense, a "bi-dual problem" to the original robust optimization one. Assumption 6 corresponds in that context to a weak constraint qualification condition of geometric (rather than topological) type.

Our general results stated later on will also cover the case of uncertainty set $\mathcal{Q}$ defined by observables h_λ taking values in vector spaces of arbitrary dimension (and with general convex subsets C_λ in each of them instead of the intervals $[a_\lambda, b_\lambda]$). We point out that the problem (2.2.29) will be solved by considering first an extension in some abstract functional space, and showing that its solution actually is in $L_{U^{-1}\circ|\cdot|}$. A characterization of the solution pair to the primal-extended dual problems will also be provided.

In checking the condition in Assumption 6, the next result (following from Léonard [2010] as explained later on), is useful:

Lemma 2.2.1 *For all $y > 0$ and $(t, s) = ((t_\lambda)_{\lambda\in\Lambda}, s) \in \mathbb{R}^\Lambda \times \mathbb{R}$ one has*

$$\nu_y(t, s) = \inf \mathbb{E}\left[ZV\left(\frac{y}{Z}\right)\right],$$

where the infimum is taken over $\{Z \in L_{|\cdot|V\circ 1/|\cdot|} : (\mathbb{E}(Zh_\lambda)_{\lambda\in\Lambda}, \mathbb{E}(Z)) = (t, s)\}$.

Notice that if the uncertainty set is determined by the expectations of finitely many observables in $\mathbb{R}$, say $n \in \mathbb{N}$ of them, the maximization problems in (2.2.29) are stated in the $n + 1$ dimensional euclidean space.

Example 2.2.3 *Consider the Samuelson model under the risk neutral measure. That is, we assume that under the reference measure $\mathbb{P}$ the price process is given by $S_t = \exp\left\{-\frac{\sigma^2}{2}t + \sigma W_t\right\}$ for some standard Brownian motion W, where $\sigma^2 > 0$ and $S_0 = 1$ (for simplicity). For $A > 0$, we consider the uncertainty set $\mathcal{Q}_A := \{\mathbb{Q} \in \mathcal{P} : \mathbb{Q} \ll \mathbb{P}, \mathbb{E}^{\mathbb{Q}}(S_T) \geq A\}$ corresponding to the one in Example 2.2.1 with $h := S_T$, and the utility function $U(x) = 2x^{1/2}$ in Example 2.2.2 with $\alpha = 1/2$. Since $S_T \in L_{U^{-1}\circ|\cdot|} = L^2$, $\frac{\mathrm{d}\mathcal{Q}_A}{\mathrm{d}\mathbb{P}} \cap L_{|\cdot|V\circ 1/|\cdot|}$ is weakly closed in $L_{|\cdot|V\circ 1/|\cdot|} = L^2$. With Girsanov Theorem we easily see that for each $A > 0$, there is a probability measure $\mathbb{Q}_A$ with $\frac{\mathrm{d}\mathbb{Q}_A}{\mathrm{d}\mathbb{P}} \in L^2$ such that*

$\mathbb{E}^{\mathbb{Q}_A}(S_T) = A$. In particular, $\mathcal{Q}_A \neq \emptyset$. Moreover, $\mathcal{Q}_A$ is closed under infinite convex combinations.

In order to check Assumption 6, notice first that for any $(a, b) \in \mathbb{R}^2_+$ with $a, b \neq 0$ there is an element $Z \in L^2, Z \geq 0$ such that $(\mathbb{E}(Z), \mathbb{E}(ZS_T)) = (a, b)$ (take e.g. $Z := a\frac{\mathrm{d}\mathbb{Q}_A}{\mathrm{d}\mathbb{P}} \in L^2$ with $\mathbb{Q}_A$ as above and $A = \frac{b}{a}$). From Lemma 2.2.1 we get aff $(\mathrm{dom}\, \nu_y)) = \mathbb{R}^2$ and from the previous we actually obtain $\mathcal{C}^\Lambda \times \{1\} \subset$ icor $(\mathrm{dom}\, \nu_y)$.

We next solve the maximization problem in (2.2.29), that is

$$\begin{aligned}\sup_{(\beta,\alpha)\in\mathbb{R}^2}\left[\inf_{c\geq A}\beta + c\alpha - \mathbb{E}^{\mathbb{P}}\left(yU^{-1}(\beta+S_T\alpha)_+\right)\right] &= \sup_{\substack{\beta\in\mathbb{R}\\ \alpha\geq 0}}\left[\beta + A\alpha - \mathbb{E}^{\mathbb{P}}\left(yU^{-1}(\beta+S_T\alpha)_+\right)\right]\\ &= \sup_{\substack{\beta\in\mathbb{R}\\ \alpha\geq 0}}\beta + A\alpha - \frac{y}{4}\mathbb{E}^{\mathbb{P}}\left((\beta+S_T\alpha)^2\mathbb{1}_{\beta+S_T\alpha>0}\right).\end{aligned} \tag{2.2.30}$$

In order to get explicit expressions, we assume in what follows that

$$e^{\sigma^2 T} > A > 1.$$

Since the function $(\beta, \alpha) \mapsto \mathbb{E}^{\mathbb{P}}\left((\beta + S_T\alpha)^2\mathbb{1}_{\beta+S_T\alpha>0}\right)$ is convex on the whole plane $\mathbb{R}^2$ (as inherited out of the convexity of U^{-1}), the function $f(\beta, \alpha)$ being maximized in the last supremum in (2.2.30) is concave on $\mathbb{R}^2$ and admits a global maximum. Since $\mathbb{E}^{\mathbb{P}}\left(S_T^2\right) = e^{\sigma^2 T}$ and $\mathbb{E}^{\mathbb{P}}(S_T) = 1$, in $\{(\beta, \alpha) \in \mathbb{R}^2 : \beta > 0, \alpha > 0\}$ we have

$$f(\beta,\alpha) = \beta + A\alpha - \frac{y}{4}\left(\beta^2 + 2\beta\alpha + e^{\sigma^2 T}\alpha^2\right),$$

whence f is twice continuously differentiable on such part of the plane. Since $\exp\{\sigma^2 T\} > A > 1$, it is verified that $(\beta^*, \alpha^*) = (\beta^*(y), \alpha^*(y)) := \left(\frac{2(e^{\sigma^2 T}-A)}{y\left(e^{\sigma^2 T}-1\right)}, \frac{2(A-1)}{y\left(e^{\sigma^2 T}-1\right)}\right)$ satisfies $(\beta^*, \alpha^*) \in (0, \infty)^2$ and $\nabla f(\beta^*, \alpha^*) = 0$. Thus (β^*, α^*) is a local maximum of f and hence also a global one. This shows, after some computations, that (2.2.30) is equal to $\frac{1}{y}\left[1 + \frac{(A-1)^2}{e^{\sigma^2 T}-1}\right]$. We obtain:

$$u(x) = \inf_{y>0}\left\{xy + \frac{1}{y}\left[1 + \frac{(A-1)^2}{e^{\sigma^2 T}-1}\right]\right\} = \left\{x\hat{y} + \frac{1}{\hat{y}}\left[1 + \frac{(A-1)^2}{e^{\sigma^2 T}-1}\right]\right\},$$

for $\hat{y} = \sqrt{\frac{1+\frac{(A-1)^2}{(e^{\sigma^2 T}-1)}}{x}}$. That is,

$$u(x) = 2\sqrt{x\left(1 + \frac{(A-1)^2}{e^{\sigma^2 T}-1}\right)}.$$

We conclude that the optimal measure is given in terms of the pair $(\beta^*, \alpha^*) = (\beta^*(\hat{y}), \alpha^*(\hat{y}))$

by $\hat{\mathbb{Q}} = \hat{y}[U^{-1}]'\left(\langle(\beta^*, \alpha^*), (1, S_T)\rangle\right)\mathbb{P}(\mathrm{d}\omega)$*, that is*

$$\hat{\mathbb{Q}}(\mathrm{d}\omega) := \frac{e^{\sigma^2 T} - A + S_T(A-1)}{e^{\sigma^2 T} - 1}\mathbb{P}(\mathrm{d}\omega).$$

Let us remark that $\hat{\mathbb{Q}}$ *is the unique convex combination of the measures* $\mathrm{d}\mathbb{P}$ *and* $S_T\mathrm{d}\mathbb{P}$ *being a probability measure and satisfying* $\mathbb{E}^{\hat{\mathbb{Q}}}(S_T) = A$*.*

Last, part b) of Theorem 2.2.5 implies that the terminal wealth of the optimal portfolio is given $\mathbb{P}$ *and* $\hat{\mathbb{Q}}$ *a.s. by*

$$\hat{X}_T := x\frac{\left(e^{\sigma^2 T} - A + S_T(A-1)\right)^2}{\left(e^{\sigma^2 T} - 1 + (A-1)^2\right)\left(e^{\sigma^2 T} - 1\right)}.$$

The robust optimal strategy can then be derived by standard hedging arguments, using the fact that $\hat{X}_T$ *is under* $\mathbb{P}$ *the final value of a martingale issued from* x *(which also follows from* $\hat{X}_t$ *being a submartingale with* $\mathbb{E}^{\mathbb{P}}(\hat{X}_T) = x$*).*

2.3 Orlicz-Musielak spaces and the robust optimization problem

We now introduce some general functional spaces needed in our study of the robust optimization problem. These can actually be seen as Orlicz spaces based on "randomized Young functions." Their main properties including dual spaces and reflexivity are first recalled, following succinctly the presentation in Kozek [1976/77] and Kozek [1980].

We then translate these concepts to the robust optimization setting, for which some relevant functionals are introduced, and their main properties are studied. The goal is to identify, from the ingredients of the financial problem, who the key "randomized Young functions" are and from them construct the Orlicz-Musielak spaces that shall be relevant in our analysis (see Remark 2.3.3 for this).

2.3.1 Orlicz-Musielak spaces

Recall that $(\Omega, \mathcal{F}, \mathbb{P})$ is a (complete) probability space and that the notation $\mathbb{E}(\cdot)$ is employed for the expectation under $\mathbb{P}$. Let us define what we meant with randomized Young functions:

Definition 2.3.1 *A functional* $\rho : (-\infty, \infty) \times \Omega \to [0, \infty]$ *is said to be a rho-functional if the following hold:*

1. ρ *is jointly measurable*
2. *for almost every* $\omega \in \Omega$*,* $\rho(\cdot, \omega)$ *is lower-semicontinuous and convex*
3. $\rho(0, \cdot) \equiv 0$ *and* $\rho(x, \cdot) = \rho(-x, \cdot)$

4. *If* $\alpha : \Omega \to (0,\infty)$ *is measurable, then there exists a measurable function* $\lambda : \Omega \to (0,\infty)$ *such that a.s.* $[|x| \geq \lambda(\omega) \Rightarrow \rho(x,\omega) \geq \alpha(\omega)]$ *.*

5. *If* $\epsilon : \Omega \to (0,\infty)$ *is measurable, then there exists a measurable function* $\rho : \Omega \to (0,\infty)$ *such that a.s.* $[|x| \leq \rho(\omega) \Rightarrow \rho(x,\omega) \leq \epsilon(\omega)]$ *.*

6. *The random variables* $\rho(x,\cdot)$ *and* $\rho^*(y,\cdot) := \sup_{x\in(-\infty,\infty)}(xy - \rho(x,\cdot))$ *are integrable for every* $x, y \in (-\infty,\infty)$*.*

Remark 2.3.1 *Under the conditions in Definition 2.3.1, the results in Kozek [1976/77] are valid. It is worth noting that in that paper a functional* ρ *satisfying conditions 1. through 5. was called an "N-function." However, such a* ρ *"only" converges a.s. to zero (resp. to* ∞*) when* x *tends to zero (resp. to* ∞*), whereas in the standard definition of N-functions, it is the quotient* $\frac{\rho(x,\omega)}{x}$ *that has this limiting behaviour in* x *near* 0 *and* $+\infty$*. To avoid confusions we use here the different "rho-functional" terminology. Also, in the language of Kozek [1976/77], the above condition 6. amounts to requiring "condition B on* ρ *and* ρ^**," and is necessary to obtain nice topological properties (see below). Last, it is not difficult to see from the above definition that* ρ^* *is also a rho-functional.*

Define now for a random variable $Z : \Omega \to (-\infty,\infty)$,

$$I_\rho(Z) := \mathbb{E}\left[\rho(Z,\cdot)\right] \leq \infty.$$

In the terminology of Kozek [1976/77], this is a normal convex modular. This allows us to define the following spaces:

Definition 2.3.2 *The Orlicz-Musielak space (or generalized Orlicz space) associated to* ρ *is defined as:*

$$L_\rho(\Omega,\mathbb{P}) := \left\{Z \in L^0 \ \text{s.t. } \exists\alpha > 0, I_\rho(\alpha Z) < \infty\right\}, \tag{2.3.1}$$

and its so-called Orlicz heart is the subspace:

$$E_\rho(\Omega,\mathbb{P}) := \left\{Z \in L^0 \ \text{s.t. } \forall\alpha > 0, I_\rho(\alpha Z) < \infty\right\}. \tag{2.3.2}$$

In the following, L_ρ will stand as an abbreviation for $L_\rho(\Omega,\mathbb{P})$. We have:

Theorem 2.3.1 *The following functionals define equivalent norms on* L_ρ*:*

$$\|Z\|_\rho^\ell := \inf\left\{\beta > 0 : I_\rho\left(\frac{Z}{\beta}\right) \leq 1\right\}, \tag{2.3.3}$$

$$\|Z\|_\rho^a := \sup\left\{\mathbb{E}(\phi Z) \,:\, \phi \in L_{\rho^*}, \hat{I}_\rho(\phi) \leq 1\right\} \tag{2.3.4}$$

$$= \sup\left\{\mathbb{E}(\phi Z) \,:\, \phi \in L_{\rho^*}, ||\phi||_{\rho^*}^\ell \leq 1\right\}, \tag{2.3.5}$$

where $\hat{I}_\rho(\phi) := \sup_{Z \in L_\rho} [\mathbb{E}(\phi Z) - I_\rho(Z)] = I_{\rho^*}$*, and* $\rho^*(\cdot,\omega)$ *is the a.s. convex conjugate of* $\rho(\cdot,\omega)$ *as defined previously. Moreover, the norm* $\|\cdot\|^a_\rho$ *has the equivalent expression*

$$\|Z\|^a_\rho = \inf_{k>0} \left\{ \frac{1}{k} \left(1 + I_\rho(kZ)\right) \right\}. \tag{2.3.6}$$

Under these equivalent norms, the linear space L_ρ *is a Banach space.*

Finally, when ρ *is finite the topological dual of* E_ρ *is isometrically isomorphic to* L_{ρ^*} *(assuming that in one space a* $\|\cdot\|^\ell$ *norm is taken and in the other a* $\|\cdot\|^a$ *norm is taken) with the identification* $[\phi \in E^*_\rho \leftrightarrow g \in L_{\rho^*}] \iff [\phi(Z) = \mathbb{E}(Zg), \forall Z \in E_\rho]$*.*

Proof. The first, second and third assertions follow from [Kozek, 1980, Propositions 1.5 and 1.6], plus [Kozek, 1976/77, Proposition 4.5 and Theorem 2.4]. The last assertion stems on the one hand from [Kozek, 1976/77, Theorem 4.8 and Proposition 3.3] (stating that the topological dual of the closure M_ρ under $\|\cdot\|_\rho$ of the linear span of simple functions is always isometrically isomorphic to L_{ρ^*}) and, on the other hand, from [Musielak, 1983, Theorem 7.6] (implying that $E_\rho = M_\rho$ when point 6. in the above definition of rho-functionals holds). ■

The norms $\|\cdot\|^\ell_\rho$ and $\|\cdot\|^a_\rho$ are called respectively Luxemburg and Amemiya norms. Now thanks to Young's inequality, one can derive a series of Hölder inequalities:

$$\mathbb{E}(|Zg|) \leq 2N_\rho(Z)N_{\rho^*}(g),$$

where N_ρ (resp. N_{ρ^*}) represents any of the norms in L_ρ (resp. L_{ρ^*}) introduced in Theorem 2.3.1. In particular, L_{ρ^*} (resp. L_ρ) is embedded in the topological dual of L_ρ (resp. L^*_ρ), and L_ρ and L_{ρ^*} are continuously embedded in L^1. The following growth property of a rho-functional and its relation with topological properties of the associated Orlicz-Musielak space is relevant:

Definition 2.3.3 *A finite rho-functional is said to satisfy the* Δ_2 *condition (or* $\rho \in \Delta_2$*), if there is a constant* $K \geq 1$ *and a non-negative integrable function* h *such that a.s.:*

$$\rho(2x,\omega) \leq K\rho(x,\omega) + h(\omega). \tag{2.3.7}$$

We then have by [Kozek, 1980, Corollary 1.7.4] that:

Theorem 2.3.2 *Let* ρ *satisfy condition* Δ_2*. Then* $E_\rho = dom\,(I_\rho) = L_\rho$ *and hence* $(L_\rho)^*$ *is isometrically isomorphic to* L_{ρ^*}*. Moreover, if the measure* $\mathbb{P}$ *is non-atomic, the condition* Δ_2 *is also necessary for this last isomorphism to hold.*

Therefore, if both ρ *and* ρ^* *satisfy the* Δ_2 *condition, the Banach spaces* L_ρ *and* L_{ρ^*} *are in topological duality and are reflexive. The converse is true if* $\mathbb{P}$ *is non-atomic.*

2.3.2 Towards the robust optimization problem

Our next aim is to associate a family of Orlicz-Musielak spaces of the previous type with the robust maximization problem:

$$u(x) := \sup_{X \in \mathcal{X}(x)} \inf_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}^{\mathbb{Q}} \left(U \left(X_T \right) \right).$$

We recall first some useful and well known properties of the function V in (2.2.5) that follow from Assumption 2:

Lemma 2.3.1 *The function V is strictly convex, l.s.c. finite and differentiable (on $(0, \infty)$), strictly decreasing, strictly positive, and satisfies:*

$$\lim_{x \to \infty} \frac{V(x)}{x} = \inf\{x : U(x) > -\infty\}, \tag{2.3.8}$$

$$V(0) = \lim_{x \to \infty} U(x). \tag{2.3.9}$$

Moreover, if U satisfies $AE(U) < 1$, then condition (2.2.23) *holds for V.*

Proof. The facts that V is strictly convex, l.s.c. finite and differentiable (on $(0, \infty)$) follow form the properties of U and standard results on Fenchel conjugates (see [Rockafellar, 1970b, Theorem 26.3]) as also does the fact that $V' = -[U']^{-1}$. Noting that $[U']^{-1}(\cdot) > 0$, V has to be strictly decreasing. By definition $V(y) \geq U(0+) = 0$, and this plus its strict decreasing character imply its strict positivity. Results (2.3.8), (2.3.9) and the last statement appear in [Gundel, 2006, Lemma 2.1.6]. ∎

The functions that are next introduced will play a central role in the sequel:

Definition 2.3.4 *For $l \geq 0$ we define the function*

$$\gamma_l^*(z) = \begin{cases} \infty & \text{if } z < 0, \\ zV\left(\frac{l}{z}\right) & \text{if } z \geq 0, \end{cases} \tag{2.3.10}$$

and we call γ_l its conjugate:

$$\gamma_l(x) = \sup_{z \geq 0} \{xz - \gamma_l^*(z)\}.$$

In robust optimization on finite-dimensional spaces, one would call this function γ_l^* the *adjoint* of V (see e.g. Ben-Tal et al. [1991]).

The next three results are probably known and certainly follow from elementary arguments. The proofs of the first two are contained in Backhoff; we show them here for completeness.

Lemma 2.3.2 *Under Assumption 2, we have*

- *The function $(y, z) \mapsto \gamma_y^*(z)$ is convex on $(0, \infty)^2$.*

- *The function $\gamma_l^*(\cdot)$ is l.s.c, strictly convex in its domain, finite, increasing and strictly positive on the positive half-line, $\gamma_l^*(0) = 0$ and $\lim_{t\to+\infty} \frac{\gamma_l^*(t)}{t} = +\infty$.*
- *The function γ_l is finite, everywhere differentiable, non-negative, not identically null and satisfies $\gamma_l(x) = 0$ if $x \le 0$. Furthermore, $\forall l > 0 : \gamma_l(\cdot) = l\gamma_1(\cdot)$.*

Proof. The first point is known (see Föllmer and Gundel [2006] or Schied and Wu [2005]): for $y_0, y_1, z_0, z_1 > 0$ and $\beta \in (0,1)$, the numbers $y_\beta := \beta y_1 + (1-\beta)y_0$, $z_\beta := \beta z_1 + (1-\beta)z_0$ and $\alpha := (\beta z_1/z_\beta) \in (0,1)$ satisfy

$$
\begin{aligned}
z_\beta V\left(\frac{y_\beta}{z_\beta}\right) =& z_\beta V(\alpha(y_1/z_1) + (1-\alpha)(y_0/z_0)) \\
&\le z_\beta(\alpha V(y_1/z_1) + (1-\alpha)V(y_0/z_0)) = \beta z_1 V(y_1/z_1) + (1-\beta)z_0 V(y_0/z_0)
\end{aligned}
$$

by convexity of V, with strict inequality if $\frac{y_0}{z_0} \neq \frac{y_1}{z_1}$. The latter holds in particular when $y_0 = y_1$ and $z_0 \neq z_1$ yielding the strict convexity in the second point. By Assumption 2, the limit in (2.3.8) equals 0 and since V is finite in the positive half-line, we conclude that γ_l^* is continuous in $[0,\infty)$ hence l.s.c. in $(-\infty,\infty)$. Strict positivity in $(0,\infty)$ is also inherited from V, and (2.3.9) together with Assumption 2 yield that $\lim_{t\to+\infty} \frac{\gamma_l^*(t)}{t} = +\infty$. The latter implies on the other hand that the recession function of γ_l^* is identically infinite, this is, γ_l^* is co-finite in the sense of [Rockafellar, 1970b, Corollary 13.3.1], which according to this same result is equivalent to γ_l being finite. Moreover, from [Rockafellar, 1970b, Theorem 26.3], convexity of $\gamma_l^*(\cdot)$ plus strict convexity in its domain (which implies that this function be essentially strictly convex), imply through this result that $\gamma_l(\cdot)$ be essentially smooth. As this last function is finite, this entails it is everywhere differentiable. Next, from identity $\gamma_l(x) = [\gamma_l^*]^* = \sup_{y>0}[xy - yV(l/y)]$ it follows that $\gamma_l(x) \ge [-\gamma_l^*(0)] = 0$, i.e. γ_l is non-negative. Moreover, $\gamma_l(0) = \sup_y[-\gamma_l^*(y)] \le 0$, from where $\gamma_l(0) = 0$, and if $x < 0$, $\gamma_l(x) = \sup_{y\ge0}[xy - \gamma_l^*(y)] \le 0$. Notice γ_l can't be null, because if it were so, γ_l^* would have some infinite value, which is a contradiction. Finally, $\sup_{y>0}[xy - yV(l/y)] = l\sup_{z>0}[xz - zV(1/z)] = l\gamma_1(x)$. ∎

Since the functions γ_l and γ_l^* take respectively the values 0 and $+\infty$ over the negative reals, it will be convenient to consider their even versions. Set

$$\bar\gamma_l(\cdot) := \max\{\gamma_l(\cdot), \gamma_l(-\cdot)\} = \gamma_l(|\cdot|).$$

Lemma 2.3.3 *Under Assumption 2, it holds for all $l > 0$ that*

$$\gamma_l^*(|\cdot|) = (\bar\gamma_l)^*(\cdot) \le \gamma_l^*(\cdot).$$

Moreover, $\gamma_l^*(|\cdot|)$ is l.s.c, strictly convex, finite and strictly positive except at $z = 0$ where it vanishes and such that $\lim_{|t|\to+\infty} \frac{\gamma_l^*(|t|)}{|t|} = +\infty$. Finally, the function $\bar\gamma_l$ is finite, everywhere differentiable, non-negative, not identically null, vanishing at $z = 0$, and satisfies $\forall l > 0, \bar\gamma_l(\cdot) = l\bar\gamma_1(\cdot)$.

Proof. Since $\gamma_l(|\cdot|) \ge \gamma_l(\cdot)$, then $(\bar\gamma_l)^*(\cdot) \le \gamma_l^*(\cdot)$. Also $(\bar\gamma_l)^*(y) = \sup_x\{xy - \gamma_l(|x|)\} = \sup_{x>0}\{x|y| - \gamma_l(x)\}$. Hence if $y > 0$, $\sup_{x>0}\{xy - \gamma_l(x)\} = (\bar\gamma_l)^*(y) \le$

$\gamma_l^*(y) = \max\{\sup_{x>0}\{xy - \gamma_l(x)\}, \sup_{x\leq 0}\{xy - \gamma_l(x)\}\}$. Now, it will be proved that $\sup_{x>0}\{xy - \gamma_l(x)\} \geq \sup_{x\leq 0}\{xy - \gamma_l(x)\} = \sup_{x\leq 0}\{xy\}$. For this, given $c \leq 0$ it will be proved that $\exists z > 0$ st. $cy \leq zy - \gamma_l(z)$, i.e. that $\gamma_l(z) \leq (z-c)y$. Notice $\gamma_l(\cdot)$ under the assumption is continuous. Hence $\exists a_0 > 0$ st. $\forall 0 < a \leq a_0$, $\gamma_l(a) \leq y$. Fix now $0 < a \leq \min\{a_0, 1\}$, y, $0 < x \leq \min\{1, \frac{c}{a-1}\}$. By convexity, follows that $\gamma_l(ax) = \gamma_l(ax + 0(1-x)) \leq x\gamma_l(a) \leq xy$. Yet since $x \leq \frac{c}{a-1}$, then $x \leq ax - c$, from where $xy \leq (ax-c)y$, and thus $\gamma_l(ax) \leq (ax-c)y$. Hence taking $z = ax > 0$, it's been shown that $y > 0$, $(\bar{\gamma}_l)^*(y) = \gamma_l^*(y)$. But $(\bar{\gamma}_l)^*$ is even, as follows from the beginning of the proof. The two last claims are simple consequence of the properties of γ_l^* and γ_l. ■

The explicit form of $\bar{\gamma}_l$ turns out to be very simple, and we shall profit from it:

Lemma 2.3.4 *The conjugate function of $\gamma_l^*(|\cdot|)$ is $\bar{\gamma}_l(\cdot) = lU^{-1}(|\cdot|)$.*

Proof. Clearly $\bar{\gamma}_l(x) = \sup_{z\geq 0}\{|x|z - zV(l/z)\}$. The first order condition for this (assuming $z \neq 0$) is $|x| - V(l/z) + \frac{l}{z}V'(l/z) = 0$. But using that $V' = -[U']^{-1}$ one gets $|x| = U([U']^{-1}(l/z))$ or better $z = \frac{l}{U' \circ U^{-1}(|x|)}$. Therefore

$$\bar{\gamma}_l(x) = \frac{|x|l}{U' \circ U^{-1}(|x|)} - \frac{l}{U' \circ U^{-1}(|x|)} V \circ U' \circ U^{-1}(|x|).$$

Using again the identity $V(y) = U([U']^{-1}(y)) - y[U']^{-1}(y)$ one arrives at $\bar{\gamma}_l(x) = lU^{-1}(|x|)$. By Lemma 2.3.2 one knows that $\bar{\gamma}_l \geq 0$ and is null only at the origin. Thus if the supremum defining it were attained at 0, since $0V(l/0) = 0$, this shows x must be null. But also $U^{-1}(0) = 0$. Hence, the asserted expression for $\bar{\gamma}_l$ is always valid. ■

Remark 2.3.2 *Note that for every $\mathbb{Q} \in \mathcal{Q}_e$, we have $\mathcal{Y}_{\mathbb{Q}}(y) = \left\{\frac{yY}{Z^{\mathbb{Q}}} : Y \in \mathcal{Y}_{\mathbb{P}}(1)\right\}$, where $Z^{\mathbb{Q}}$ is the density process of $\mathbb{Q}$ w.r.t. $\mathbb{P}$. Thus we obtain:*

$$v(y) = \inf_{\mathbb{Q}\in\mathcal{Q}_e} \inf_{Y\in\mathcal{Y}_{\mathbb{P}}(1)} \mathbb{E}^{\mathbb{P}}\left[\frac{d\mathbb{Q}}{d\mathbb{P}} V\left(yY_T\left[\frac{d\mathbb{Q}}{d\mathbb{P}}\right]^{-1}\right)\right]. \tag{2.3.11}$$

This implies that if v is to be finite at some point $y > 0$, the only measures $\mathbb{Q}$ that matter in (2.3.11) *are those such that, for some $Y \in \mathcal{Y}_{\mathbb{P}}(1)$,*

$$\mathbb{E}^{\mathbb{P}}\left[\frac{d\mathbb{Q}}{d\mathbb{P}} V\left(yY_T\left[\frac{d\mathbb{Q}}{d\mathbb{P}}\right]^{-1}\right)\right] < \infty.$$

Since those Y_T vanishing on a set of positive measure would induce the expectation in (2.3.11) *to be equal to $+\infty$ (since $V(0) = U(+\infty) = +\infty$ by assumption), there is no loss of generality in considering only almost surely strictly positive Y_T when studying* (2.3.11). *Notice that strictly positive elements in $\mathcal{Y}_{\mathbb{P}}(1)$ do exist, see e.g. [Pratelli, 2005, Lemma 4.1].*

This leads us to introduce

Definition 2.3.5 *Let $Y \in \mathcal{Y}_{\mathbb{P}}(1)$. We denote by $\eta_Y^*, \eta_Y : (-\infty,\infty) \times \Omega \to [0,\infty]$ the functionals respectively given by*

$$\eta_Y^*(z,\omega) := \gamma_{Y_T(\omega)}^*(|z|) = |z| V\left(\frac{Y_T(\omega)}{|z|}\right),$$

and

$$\eta_Y(z,\omega) := \gamma_{Y_T(\omega)}(|z|) = Y_T(\omega) U^{-1}(|z|).$$

Of course, if $Y_T > 0$ a.s., $\eta_Y^*(\cdot,\omega)$ and $\eta_Y(\cdot,\omega)$ almost surely inherit the properties of $\gamma_l^*(|\cdot|)$ and $\gamma_l(|\cdot|)$ stated in Lemma 2.3.3. As it is next proved, under mild assumptions they induce rho-functionals.

Proposition 2.3.1 *Let $Y \in \mathcal{Y}_{\mathbb{P}}(1)$ be strictly positive a.s. and suppose Assumption 2.*

a) Then, a.s. the convex conjugate of the function $\eta_Y^*(\cdot,\omega)$ is $\eta_Y(\cdot,\omega)$ and, provided that

$$\forall \beta > 0, \mathbb{E}[V(\beta Y_T)] < \infty,$$

$\eta_Y^*(\cdot,,\omega)$ and $\eta_Y(\cdot,,\omega)$ are rho-functionals in the sense of Definition 2.3.1.

b) If condition (2.2.23) (resp (2.2.24)) holds, the function $\eta_Y^*(\cdot,,\omega)$ (resp. $\eta_Y(\cdot,,\omega)$) is in Δ_2.

c) If $AE(U) < 1$, then $\eta_Y^* \in \Delta_2$ and the condition in a) reduces to

$$\exists \beta > 0, \mathbb{E}[V(\beta Y_T)] < \infty.$$

Proof. The functionals η_Y and η_Y^* are clearly jointly measurable, and the fact that they are conjugate to each other follows from applying Lemma 2.3.4 almost surely. By properties of U and V, as functions of z they are a.s. l.s.c., even, null at the origin and convergent to 0 at 0 and to infinity at infinity. Also, $\mathbb{E}[Y_T U^{-1}(c)] \leq U^{-1}(c)$ for every constant $c > 0$ since $Y \in \mathcal{Y}_{\mathbb{P}}(1)$ satisfies $\mathbb{E}(Y_T) \leq 1$. Hence, $\eta_Y(c)$ is integrable. The assumption $\mathbb{E}[V(\beta Y_T)] < \infty$ for every $\beta > 0$ implies that also η_Y^* is integrable when applied to constants. We conclude that they are rho-functional. For the second point, notice that thanks to (2.2.23),

$$\begin{aligned}
\eta_Y^*(2z) =& 2zV\left(\frac{Y}{2z}\right) \leq 2a\eta_Y^*(z) + 2b(Y+z) \\
=& 2a\eta_Y^*(z) + 2bY + 2bz\mathbf{1}_{\{z \geq Y/V^{-1}(1)\}} + 2bz\mathbf{1}_{\{z < Y/V^{-1}(1)\}} \\
\leq& 2a\eta_Y^*(z) + 2bY + 2b\eta_Y^*(z) + 2bY/V^{-1}(1),
\end{aligned}$$

for every $z > 0$, which means that $\eta_Y^* \in \Delta_2$. The corresponding property for η_Y is direct. The last statement c) follows from the last part of Lemma 2.3.1. ∎

Point (c) above should be compared with the comment before [Kramkov and Schachermayer, 1999, Corollary 6.1].

With some abuse of notation, for $Z \in L^0$ we will write simply $\eta_Y^*(Z)$ referring to the function $\eta_Y^*(Z,\cdot) : \Omega \to [0,+\infty)$ such that $\eta_Y^*(Z,\cdot)(\omega) = \eta_Y^*(Z(\omega),\omega)$.

Remark 2.3.3 *We deduce that if $Y \in \mathcal{Y}_{\mathbb{P}}(1)$ satisfies $Y_T > 0$ a.s. and $\mathbb{E}[V(\beta Y_T)] < \infty, \forall \beta > 0$, the space*

$$L_{\eta_Y^*} = \left\{ Z \in L^0 \text{ s.t. } \exists \alpha > 0, \mathbb{E}^{\mathbb{P}} \left[\eta_Y^* \left(\alpha Z \right) \right] < \infty \right\},$$

is an Orlicz-Musielak space. Moreover, $L_{\eta_Y^*}$ and L_{η_Y} (defined analogously) are in separating topological duality and by [Kozek, 1976/77, Theorem A.5] or [Kozek, 1980, Proposition 1.5] we get that $\mathbb{E}[\eta_Y^*(\cdot)]$ and $\mathbb{E}[\eta_Y(\cdot)]$ are convex conjugates to each other w.r.t. the given duality.

Now some enlightening topological aspects of $L_{\eta_Y^*}$ are given:

Lemma 2.3.5 *Assume that for all $\beta > 0$, $\mathbb{E}[V(\beta Y_T)] < \infty$.*

- *If $\{Z_n\} \subset L_{\eta_Y^*}$ converges to 0, so it does in L^1.*
- *Bounded subsets of $L_{\eta_Y^*}$ are uniformly integrable.*

Proof. Since η_Y^* is a rho-functional, Hölder's inequality implying that $L_{\eta_Y^*}$ in injected continuously in L^1 yields the first point. For the second point, let $\mathcal{K}$ be a bounded subset of $L_{\eta_Y^*}$. Since it is bounded in L^1 by the previous point, we only need to show that $\forall \epsilon > 0, \exists \delta > 0$ such that $\mathbb{P}(A) \leq \delta$ implies $\forall A \in \mathcal{K} : \int_A Z \mathrm{d}\mathbb{P} < \epsilon$. So first fix $\epsilon > 0$. From the aforementioned Hölder inequality for Orlicz-Musielak spaces, we have that $\mathbb{E}\left[Z \mathbb{1}_A\right] \leq 2\|Z\|_{\eta_Y^*} \|\mathbb{1}_A\|_{\eta_Y} \leq c \|\mathbb{1}_A\|_{\eta_Y}$. Now, notice that $\forall l > 0$, $\|\mathbb{1}_A\|^a_{\eta_Y} \leq \frac{1}{l}\left[1 + U^{-1}(l) \int_A Y_T \mathrm{d}\mathbb{P}\right]$, since $U^{-1}(0) = 0$. Since Y_T is integrable $L := \int_A Y_T \mathrm{d}\mathbb{P}$ can be made arbitrarily small as the measure of A tends to zero. Taking $l = U\left(\frac{1}{L}\right)$ yields $\|\mathbb{1}_A\|^a_{\eta_Y} \leq \frac{2}{U\left(\frac{1}{L}\right)}$, from where we conclude that $\mathbb{E}\left[Z \mathbb{1}_A\right] \leq \epsilon$ if $\mathbb{P}(A)$ is small enough. ■

In the next section we will apply the Orlicz-Musielak point of view in detail in the case of a complete market, in order to get rid of the assumption of closedness in L^0 of the set $\frac{\mathrm{d}\mathcal{Q}}{\mathrm{d}\mathbb{P}}$, and thus extend some of the results of Föllmer and Gundel [2006] and Schied and Wu [2005].

2.4 The Complete case

For notational simplicity we assume that the reference measure is the unique martingale measure. The results can be readily generalized if this were not the case, at the price of dealing with the random Young functions η_Y^*, η_Y (where Y is the density of the unique martingale measure) instead of the deterministic ones that we will encounter. Under this assumption, [Kramkov and Schachermayer, 1999, Lemma 4.3] and its proof states that every terminal value of the elements $Y \in \mathcal{Y}_{\mathbb{P}}(1)$ is bounded by 1 and (since V is non-increasing) we have:

$$v(y) = \inf_{Z \in \frac{\mathrm{d}\mathcal{Q}_e}{\mathrm{d}\mathbb{P}}} \mathbb{E}\left[ZV\left(\frac{y}{Z}\right)\right]. \tag{2.4.1}$$

The only Orlicz-Musielak space pertinent to the problem is thus the Orlicz space

$$L_{\eta^*} = \left\{Z \in L^0 \text{ s.t. } \exists \alpha > 0, \mathbb{E}^{\mathbb{P}}\left[\eta^*(\alpha Z)\right] < \infty\right\},$$

associated with

$$\eta^*(z) := \eta_1^*(z) = |z| V\left(\frac{1}{|z|}\right) = \gamma_1^*(|z|),\ z \in (-\infty, \infty). \tag{2.4.2}$$

Recall from Lemma 2.3.4 that the conjugate function of η^* is the even function

$$\eta := \bar{\gamma}_1(\cdot) = \gamma_1(|\cdot|) = U^{-1}(|\cdot|).$$

In the results to be established in this section, the plain idea is to recover and indeed sharpen the results already known in the literature (Kramkov and Schachermayer [1999] in the non-robust case and Schied and Wu [2005], Gundel [2006] in the robust case), replacing weak compactness in L^1 of $\frac{\mathrm{d}\mathcal{Q}}{\mathrm{d}\mathbb{P}}$ by weak closedness in L_{η^*} and reflexivity of that space (for some results it will be enough to have this space be a norm-dual one), for instance by characterizing the worst-case measures. Additional relevant properties of $\mathcal{Q}$ in Schied and Wu [2005] which are obtained as consequence of the L^0-closedness will be provided here by our assumptions on $\mathcal{Q}$.

We remark that norm-bounds, the minimax equality, attainability of strategies and conjugacy between u and v, shall be obtained in the incomplete market setting. For pedagogical reasons, we state without proof in this section these results, and then establish further specific results that are not covered by the incomplete-case analysis.

2.4.1 Solving the robust optimization problem without weak L^1- compactness

We will make throughout this section the assumption:

Assumption 7

- *$\mathcal{Q}$ is countably convex.*

- $[\mathbb{P}(A) = 0 \iff \forall \mathbb{Q} \in \mathcal{Q}, \mathbb{Q}(A) = 0]$.
- $\frac{\mathrm{d}\mathcal{Q}}{\mathrm{d}\mathbb{P}}$ *is a non-empty* $\sigma(L_{\eta^*}, E_\eta)$*-weakly (i.e. weak-star) closed subset of* L_{η^*}.

Remark 2.4.1 *Under the first two points above, again Halmos-Savage Theorem guarantees that* $\mathcal{Q}_e$ *is non-empty. We could have naturally assumed convexity of* $\mathcal{Q}$ *and non-triviality of* $\mathcal{Q}_e$ *only.*

The next result, is a special case of Proposition 2.5.5 valid in the incomplete case:

Proposition 2.4.1 *Suppose Assumptions 2 and 7, and moreover that* $\exists x_0 > 0, \exists \mathbb{Q}_0 \in \mathcal{Q}_e$ *such that* $u_{\mathbb{Q}_0}(x_0) < \infty$. *Then, for all* $x > 0$ *we have that*

$$\forall \mathbb{Q} \in \mathcal{Q}: \quad (1+x) \left\| \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} \right\|_{\eta^*}^{\ell} \geq u_{\mathbb{Q}}(x) \geq (1 \wedge x) \left\| \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} \right\|_{\eta^*}^{a}. \tag{2.4.3}$$

Let us also specialize Theorem 2.5.1, dealing with minimax equality and optimal strategies in the incomplete case, to the current setting:

Theorem 2.4.1 *Suppose Assumptions 2 and 7, and assume that the space* L_{η^*} *is reflexive (e.g.* $\eta^* \in \Delta_2$ *and* $\eta \in \Delta_2$*). Assume that* $\exists x_0 > 0$ *such that* $u_{\mathbb{Q}_0}(x_0) < \infty$ *for some* $\mathbb{Q}_0 \in \mathcal{Q}_e$. *Then for every* $x > 0$*:*

$$\begin{aligned} u(x) \quad &= \inf_{\mathbb{Q}\in\mathcal{Q}} \sup_{X\in\mathcal{X}(x)} \mathbb{E}^{\mathbb{Q}}\left(U\left(X_T\right)\right) \quad = \min_{\mathbb{Q}\in\mathcal{Q}} \max_{X\in\mathcal{X}(x)} \mathbb{E}^{\mathbb{Q}}\left(U\left(X_T\right)\right) \\ &= \inf_{\mathbb{Q}\in\mathcal{Q}_e} \sup_{X\in\mathcal{X}(x)} \mathbb{E}^{\mathbb{Q}}\left(U\left(X_T\right)\right) \quad < +\infty. \end{aligned} \tag{2.4.4}$$

Moreover, v *is finite and* u, v *are conjugates on* $(0,\infty)$*, and so we have:*

$$u(x) = \inf_{y>0} \inf_{\mathbb{Q}\in\mathcal{Q}_e} \{v_{\mathbb{Q}}(y) + xy\} = \inf_{y>0}\{v(y)+xy\} = \inf_{y>0} \left\{ \inf_{\mathbb{Q}\in\mathcal{Q}_e} \int \left[\gamma_y^* \left(\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} \right) \mathrm{d}\mathbb{P} \right] + xy \right\}.$$

Remark 2.4.2 *The condition* $[\exists x_0 > 0$ *such that* $u_{\mathbb{Q}_0}(x_0) < \infty$, *for some* $\mathbb{Q}_0 \in \mathcal{Q}_e]$ *has several consequences: first* $u(\cdot)$ *must be finite, second and in view of the lower bound in* (2.4.3) *and the minimax Theorem (so assuming reflexivity) we see that if* $\mathcal{Q}$ *had measures outside* L_{η^*} *then these would not count for* u *(this is the connection between Assumptions 7 and 4), and third that again by reflexivity for* $\mathbb{Q} \in \mathcal{Q}_e$ *the function* $v_{\mathbb{Q}}$ *(and hence* v*) must be everywhere finite owing to the* Δ_2 *condition and by the previous point.*

Since we will have proved the minimax equality (2.4.4), the estimates (2.4.3) (e.g. the lower bound therein) and clearly $\|Z\|_{\eta^*}^a \leq y + \mathbb{E}[|Z|V(y/|Z|)]$, we could separately reduce the problems $\inf_{\mathbb{Q}} u_{\mathbb{Q}}(x)$ and $\inf_{\mathbb{Q}} v_{\mathbb{Q}}(y)$ to subsets of $\mathcal{Q}$ whose densities become weakly-compact sets in L^1, and actually these subsets could be chosen fixed for neighbourhoods around x and y respectively. Although probably feasible, it is not obvious how to connect these local reductions with the original problem (u and v) since convex conjugacy is not simply localizable. We thus choose not to embed (locally) our problem in L^1 and instead stay in our Orlicz space, and follow the route in Schied and Wu [2005] generalizing and applying the results therein as needed.

We now prove an attainability/stability result, which we will subsequently need.

Proposition 2.4.2 *Under the same assumptions of Theorem 2.4.1, for every $x, y > 0$ there exists $\hat{Z}, Z \in \frac{\mathrm{d}\mathcal{Q}}{\mathrm{d}\mathbb{P}}$ such that (calling $\hat{\mathbb{Q}} = \hat{Z}\mathrm{d}\mathbb{P}, \mathbb{Q} = Z\mathrm{d}\mathbb{P}$),*

$$u(x) = u_{\hat{\mathbb{Q}}}(x) := \sup_{X \in \mathcal{X}(x)} \mathbb{E}^{\hat{Q}}\left(U\left(X_T\right)\right) \text{ and } v(y) = v_{\mathbb{Q}}(y) = \mathbb{E}[ZV(y/Z)]. \qquad (2.4.5)$$

Moreover, $\hat{Z}$ can be chosen to be the strong L_{η^*} limit of a sequence $\{B_n\}_n \subset \frac{\mathrm{d}\mathcal{Q}_e}{\mathrm{d}\mathbb{P}}$ such that $u(x) = \lim u_{B_n \mathrm{d}\mathbb{P}}(x)$, and Z the strong L_{η^*} limit of a sequence $\{Z_n\}_n \subset \frac{\mathrm{d}\mathcal{Q}_e}{\mathrm{d}\mathbb{P}}$ such that $\mathbb{E}[\gamma_y^*(Z_n)] := E[Z_n V(y/Z_n)] \to v(y)$.

Proof. As explained in Schied and Wu [2005], if $y > 0$ is such that $v(y) < \infty$ (which is the case, see Remark 2.4.2) then:

$$v(y) = \inf_{\mathbb{Q} \in \mathcal{Q}_e} \mathbb{E}^{\mathbb{Q}}[V(y\mathrm{d}\mathbb{P}/\mathrm{d}\mathbb{Q})] = \inf_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}^{\mathbb{Q}}[V(y[\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}]^{-1})].$$

Hence for the second statement let $W_n \in \frac{\mathrm{d}\mathcal{Q}_e}{\mathrm{d}\mathbb{P}}$ be such that

$$\mathbb{E}^{W_n \mathrm{d}\mathbb{P}}[V(yW_n^{-1})] = \mathbb{E}[W_n V(y/W_n)] \searrow v(y).$$

Due to the simple bound $\|W_n\|_{\eta^*}^a \leq y + \mathbb{E}[W_n V(y/W_n)]$ we see that the sequence is bounded and thus except for a subsequence it is weakly convergent (in $\frac{\mathrm{d}\mathcal{Q}}{\mathrm{d}\mathbb{P}}$, by assumption). Recalling Mazur's Lemma, which allows to pass from weak to strong convergence by convex combinations of the tail of the sequence, we therefore find $Z_n \to Z$ strongly and by assumption the Z_n's live in $\frac{\mathrm{d}\mathcal{Q}_e}{\mathrm{d}\mathbb{P}}$ and Z in $\frac{\mathrm{d}\mathcal{Q}}{\mathrm{d}\mathbb{P}}$. Notice $v(y) \leq \liminf \mathbb{E}[Z_n V(y/Z_n)] \leq \lim \mathbb{E}[W_n V(y/W_n)] = v(y)$, by convexity of $\mathbb{E}[\cdot V(y/\cdot)]$ plus the choice of $\{W_n\}$ and $\{Z_n\}$, from which analogously $v(y) = \lim \mathbb{E}[Z_n V(y/Z_n)]$. Building on Remark 2.3.3 we see that $\mathbb{E}[|\cdot| V(y/|\cdot|)]$ is a conjugate function and thus l.s.c., from which Z indeed attains $v(y)$.

From Theorem 2.4.1, a sequence $\{A_n\}_n \subset \frac{\mathrm{d}\mathcal{Q}_e}{\mathrm{d}\mathbb{P}}$ such that $u_{A_n}(x) \searrow u(x)$ exists. As in the previous paragraph, and out of the convexity of $Z \mapsto u_{Z\mathrm{d}\mathbb{P}}(x)$ plus the lower bound in (2.4.3), a further sequence $\{B_n\}_n \subset \frac{\mathrm{d}\mathcal{Q}_e}{\mathrm{d}\mathbb{P}}$ can be found, such that $u(x) = \lim u_{B_n}(x)$ and it is convergent strongly to a certain $\hat{Z}$. Since $Z \mapsto u_{Z\mathrm{d}\mathbb{P}}(x) = \sup_{H \in U(\mathcal{X}(x))} \mathbb{E}[ZH]$ is weakly l.s.c. in L_{η^*}, by virtue of $U(\mathcal{X}(x)) \subset L_\eta$, we see that $\hat{Z}\mathrm{d}\mathbb{P}$ attains $\inf u_{\mathbb{Q}}(x) = u(x)$. ■

We can now prove Theorem 2.2.5, as was stated in the overview cection 2.2.2. This is the main result of the present section, as it extends in the complete setting the main results in Schied and Wu [2005]. Notice that we avoid using Komlos-type arguments (see [Delbaen and Schachermayer, 1994, Lemma A.1.1]), as usually done in Schied and Wu [2005] and elsewhere, by employing instead our reflexive Orlicz spaces.

Proof. (of Theorem 2.2.5) By Remark 2.4.2 we see that Assumption 4 implies that w.l.o.g. we may suppose Assumption 7. Also, Assumption 3 implies that we are in the reflexive case. Thus part (a) in Theorem 2.2.5 is a consequence of Theorem 2.4.1.

We next recover [Schied and Wu, 2005, Lemma 4.1], following its proof closely. Fixing $x > 0$ and taking B_n and $\hat{Z}$ as in Proposition 2.4.2 we still find that any accumulation point of $u'_{B_n \mathrm{d}\mathbb{P}}(x)$ is contained in the superdifferential of u at x. Now, by the usual non-robust duality we know that

$$u(x) = \lim u_{B_n \mathrm{d}\mathbb{P}}(x) = \lim v_{B_n \mathrm{d}\mathbb{P}}(y_n) + xy_n \geq \mathbb{E}[\hat{Z}V(\hat{y}/\hat{Z})] + x\hat{y},$$

where $y_n = u'_{B_n \mathrm{d}\mathbb{P}}(x)$ and we eventually passed to a subsequence so that $y_n \to \hat{y}$ and used that $\mathbb{E}[\cdot V(1/\cdot)]$ is l.s.c. Since $\hat{y}$ is in the superdifferential of u at x, we have $u(x) = v(\hat{y}) + x\hat{y}$ and finally conclude that $v(\hat{y}) = \mathbb{E}[\hat{Z}V(\hat{y}/\hat{Z})]$.

The existence of an optimal strategy is known from the minimax Theorem. The proof of the explicit expression for $\hat{X}$ (on the support of $\hat{Z}$) then proceeds as in the proof of [Schied and Wu, 2005, Theorem 2.6].

Finally for part (c) of our theorem, we start by noticing that [Schied and Wu, 2005, Lemma 4.2] remains true (except for an adaptation to the complete case), since it does not employ the topology of $\mathcal{Q}$. This and again the proof of Theorem 2.6 imply that $X_T = 0 \iff \frac{\mathrm{d}\hat{\mathbb{Q}}}{\mathrm{d}\mathbb{P}} = 0$ ($\mathbb{P}-$as), and so the expression for X_T is valid $\mathbb{P}-$a.s., since $[U']^{-1}(\infty) = 0 = X_T$ in $\left\{\frac{\mathrm{d}\hat{\mathbb{Q}}}{\mathrm{d}\mathbb{P}} = 0\right\}$, $\mathbb{P}-$a.s. Strict differentiability of u again follows from the proof of Theorem 2.6. This finishes our proof. ■

We now attempt to characterize the worst-case measure associated to $u(x)$, by using the fact that it must minimize $v_{\mathbb{Q}}(y)$; see Theorem 2.2.5.

2.4.2 Characterization of the minimizing measure

Our goal now is to provide first an alternative (theoretically computable) expression for the value of

$$v(y) = \inf_{\mathbb{Q} \in \mathcal{Q}} \int \gamma_y^* \left(\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right) \mathrm{d}\mathbb{P}, \tag{2.4.6}$$

and then to describe the optimal $\hat{\mathbb{Q}}$ in part b) of Theorem 2.2.5, by using general entropy minimization results in Léonard [2008].

To that end, it will be convenient to embed first the minimization problem (2.4.6) in the space

$$\mathcal{M}_f = \mathcal{M}_f(\Omega, \mathcal{F}_T),$$

of finite signed measures on $(\Omega, \mathcal{F}_T)$ (endowed with the total variation norm), and to describe the uncertainty set $\mathcal{Q}$ so that the results of Léonard [2008] can be applied when possible. This will allow us to state then a general result characterizing the minimizing measure. We will finally deduce the proof of Theorem 2.2.6 as a particular application.

In what follows, Assumptions 2 and 3 are enforced, in particular we have that $L_\eta = E_\eta$ and the spaces L_η and L_{η^*} are in (reflexive) duality. We shall use the Luxemburg norms in these spaces unless otherwise stated. Assumptions on $\mathcal{Q}$ will be specified as needed.

For each $y > 0$, we set

$$\Psi_y(M) := \begin{cases} \int_\Omega \gamma_y^* \left(\frac{\mathrm{d}M}{\mathrm{d}\mathbb{P}}\right) \mathrm{d}\mathbb{P} & \text{if } M \ll \mathbb{P} \\ +\infty & \text{otherwise.} \end{cases} \tag{2.4.7}$$

Notice that $L_{\eta^*} = \left\{Z \in L^0 \text{ s.t. } \exists\alpha > 0, \mathbb{E}^{\mathbb{P}}\left[\eta^*(\alpha Z)\right] < \infty\right\}$ is continuously embedded into $\mathcal{M}_f$ through the mapping

$$Z \in L_{\eta^*} \mapsto M = Z \cdot \mathrm{d}\mathbb{P} \in \mathcal{M}_f$$

and can thus be seen as a subspace of $\mathcal{M}_f$ on which the convex function Ψ_y can be evaluated. Since $\gamma_y^* = +\infty$ in $(-\infty, 0)$, any $M \in \mathcal{M}_f$ that is feasible for (2.4.6) must have a non-negative density with respect to $\mathbb{P}$. Therefore, problem (2.4.6) is equivalent to minimizing Ψ_y over $\mathcal{M}_f$ under the constraints that $\int_\Omega \mathrm{d}M = 1$ and $M \in \mathcal{Q}$.

The description of the uncertainty set requires the following elements:

i) Let $(\mathbf{F}_0, \mathbf{G}_0)$ be a pair of linear spaces of arbitrary dimension, such that $\mathbf{F}_0$ is the algebraic dual of $\mathbf{G}_0$; we denote this by $\mathbf{F}_0 = (\mathbf{G}_0)'$ and we write $\langle \cdot, \cdot \rangle_{\mathbf{G}_0, \mathbf{F}_0}$ for the corresponding dual product.

ii) Let $\theta : \Omega \to \mathbf{F}_0$ a function.

iii) Let $\mathbf{C}_0 \subset \mathbf{F}_0$ be a convex subset.

The function θ is interpreted as an "observable" of the market taking values in $\mathbf{F}_0$. We will consider uncertainty sets $\mathcal{Q}$ characterized by distributional constraints on θ, to be expressed in terms of the set $\mathbf{C}_0$. To make this precise, recall that the function $\gamma_y = (\gamma_y^*)^*$ introduced in Lemma 2.3.2 satisfies $\gamma_y = y\gamma_1$ and that $\eta = \bar{\gamma}_1(\cdot) = \gamma_1(|\cdot|) = U^{-1}(|\cdot|)$. We will then write

$$\gamma := \gamma_1,$$

and enforce in what follows:

Assumption 8

i) $\forall g \in \mathbf{G}_0$, *the function* $\omega \in \Omega \mapsto \langle g, \theta(\omega) \rangle_{\mathbf{G}_0, \mathbf{F}_0}$ *is measurable.*

ii) $\forall g \in \mathbf{G}_0$, $\int \eta\left(\langle g, \theta \rangle_{\mathbf{G}_0, \mathbf{F}_0}\right) \mathrm{d}\mathbb{P} < \infty$; *equivalently,* $\forall g \in \mathbf{G}_0, \langle g, \theta \rangle_{\mathbf{G}_0, \mathbf{F}_0} \in E_\eta$.

iii) $\forall (g, a) \in \mathbf{G}_0 \times (-\infty, \infty)$, *one has* $\langle g, \theta(\cdot) \rangle_{\mathbf{G}_0, \mathbf{F}_0} = a, \mathbb{P}$ *-a.s. iff* $g = 0$ *and* $a = 0$.

iv) The set $\mathcal{Q}$ *is given by*

$$\mathcal{Q} := \left\{ \mathbb{Q} \ll \mathbb{P} \textit{ probability measure s.t. } \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} \in L_{\eta^*} \textit{ and } \Theta\left(\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right) \in \mathbf{C}_0 \right\},$$

where $\Theta : L_{\eta^*} \to \mathbf{F}_0$ *is the linear operator* $\Theta(Z) = \int \theta Z \,\mathrm{d}\mathbb{P}$ *defined by*

$$\langle g, \Theta(Z) \rangle_{\mathbf{G}_0, \mathbf{F}_0} = \int_\Omega \langle g, \theta \rangle_{\mathbf{G}_0, \mathbf{F}_0} Z \,\mathrm{d}\mathbb{P},$$

for all $g \in \mathbf{G}_0$*.*

Remark 2.4.3 *Under points i) and ii) of Assumption 8, for each* $M \ll \mathbb{P}$ *with* $\frac{\mathrm{d}M}{\mathrm{d}\mathbb{P}} \in L_{\eta^*}$ *the integral*

$$\int_\Omega \langle g, \theta(\omega)\rangle_{\mathbf{G}_0,\mathbf{F}_0} M(\mathrm{d}\omega),$$

is well defined for all $g \in \mathbf{G}_0$*, since* $\int \left|\langle g,\theta\rangle_{\mathbf{G}_0,\mathbf{F}_0}\right| M(\mathrm{d}\omega) \le 2\left\|\langle g,\theta\rangle_{\mathbf{G}_0,\mathbf{F}_0}\right\|_{L_\eta} \left\|\frac{\mathrm{d}M}{\mathrm{d}\mathbb{P}}\right\|_{L_{\eta^*}}$ *by Hölder inequality. It therefore defines an element of* $\mathbf{F}_0 = (\mathbf{G}_0)'$ *denoted by* $\Theta(Z)$ *in point iv). Observe also that if* Θ^* *denotes the adjoint of* $\Theta : L_{\eta^*} \to \mathbf{F}_0$*, defined by* $\Theta^*(g)(\omega) = \langle g, \theta(\omega)\rangle_{\mathbf{G}_0,\mathbf{F}_0}$*, point i) can also be stated as* $\Theta^*(\mathbf{G}_0) \subset E_\eta$*. The use of point iii) will become clear below. It is actually not an effective restriction. Finally, notice that* $\mathcal{Q}$ *as in point iv) is always convex though not necessarily countably convex.*

We write now

$$\mathbf{F}_1 := \mathbf{F}_0 \times \mathbb{R}, \quad \mathbf{G}_1 = \mathbf{G}_0 \times \mathbb{R},$$

and notice that $\mathbf{F}_1 = (\mathbf{G}_1)'$ (algebraic dual) with the obvious duality product, denoted $\langle\cdot,\cdot\rangle_{\mathbf{G}_1,\mathbf{F}_1}$. Set also

$$\theta_1(\omega) := (\theta(\omega), 1) \in \mathbf{F}_1, \quad \Theta_1(Z) := \left(\int \theta Z \mathrm{d}\mathbb{P}, \int Z\mathrm{d}\mathbb{P}\right) = \int \theta_1 Z \mathrm{d}\mathbb{P} \in \mathbf{F}_1,$$

and

$$\mathbf{C}_1 := \mathbf{C}_0 \times \{1\}.$$

With the previous objects, under Assumption 8 the problem (2.4.6) can be written as the following (primal) convex optimization problem in $\mathcal{M}_f$ with convex constraints:

$$\textit{Minimize } \Psi_y(M), \text{ subject to } M \ll \mathbb{P}, \frac{\mathrm{d}M}{\mathrm{d}\mathbb{P}} \in L_{\eta^*} \text{ and } \Theta_1\left(\frac{\mathrm{d}M}{\mathrm{d}\mathbb{P}}\right) \in \mathbf{C}_1. \tag{PC$_y$}$$

In order to apply the results in Léonard [2008] based on Fenchel duality for the problem (PC_y), we next introduce its dual. Observe to that end that

$$\Theta_1^*(\mathbf{G}_1) = \{\langle g_0, \theta(\cdot)\rangle_{\mathbf{G}_0,\mathbf{F}_0} + a : \ g_0 \in \mathbf{G}_0, a \in \mathbb{R}\},$$

is a linear subspace of L_η, by point ii) in Assumption 8. Also, because of point iii), the linear span in $\mathbf{F}_1$ of the range of Θ_1 is in separating duality with $\mathbf{G}_1$, the function $g \in \mathbf{G}_1 \mapsto \|\Theta_1^*(g)\|_\eta$ defines a norm and $\Theta_1^* : \mathbf{G}_1 \to L_\eta$ is an injection. In particular $\mathbf{G}_1$ can be identified with $\Theta_1^*(\mathbf{G}_1)$. Point iii) can always be assumed to hold, replacing $\mathbf{G}_1$ by $\mathbf{G}_1/$ Ker Θ_1^* if needed.

Introduce now the completion $\mathbf{G}$ of $\mathbf{G}_1$ with respect to $\|\Theta_1^*(\cdot)\|_\eta$, which is isometrically isomorphic to the closure $\overline{\Theta_1^*(\mathbf{G}_1)}^{L_\eta}$ in L_η. The mapping Θ_1^* has a natural equally denoted isometric extension to $\mathbf{G}$ and, with some abuse of notation, we write

$$\langle g, \theta_1\rangle := \Theta_1^*(g) \tag{2.4.8}$$

for the element of $\overline{\Theta_1^*(\mathbf{G}_1)}^{L_\eta}$ identified with $g \in \mathbf{G}$. We also denote by $\mathbf{F}$ the topological dual $\mathbf{F} = \mathbf{G}^*$ of $\mathbf{G}$ (or $\mathbf{G}_1$, equivalently):

$$\mathbf{F} := \{f \in \mathbf{F}_1 \ : \exists C_f > 0 \text{ s.t. } |\langle g, f\rangle_{\mathbf{G}_1,\mathbf{F}_1}| \le C_f \|\Theta_1^*(g)\|_\eta \ \forall g \in \mathbf{G}_1\},$$

and we use the notation $\langle\cdot,\cdot\rangle$ also for the natural extension of the dual product $\langle\cdot,\cdot\rangle_{\mathbf{G}_1,\mathbf{F}_1}$ from $\mathbf{G}_1 \times \mathbf{F}$ to $\mathbf{G} \times \mathbf{F}$, namely $\langle g, f\rangle := \lim_{n\to\infty}\langle g_n, f\rangle_{\mathbf{G}_1,\mathbf{F}_1}$ for any sequence $g_n \in \mathbf{G}_1$ such that $\|\langle g_n, \theta_1\rangle_{\mathbf{G}_1,\mathbf{F}_1} - \langle g, \theta_1\rangle\|_\eta \to 0$.

Notice that Θ_1 continuously embeds L_{η^*} into $\mathbf{F}$. Moreover, by Hahn-Banach extension Theorem, with each $f \in \mathbf{F}$ one can associate an element $Z^f \in L_{\eta^*}$ such that

$$\langle g, f\rangle = \int Z^f \langle g, \theta_1\rangle \mathrm{d}\mathbb{P} = \langle g, \Theta_1(Z^f)\rangle \text{ for all } g \in \mathbf{G}_1. \tag{2.4.9}$$

In other words, $\Theta_1 : L_{\eta^*} \mapsto \mathbf{F}$ is surjective. Thus, $\mathbf{F}$ can be identified with the quotient of L_{η^*} by the annihilator $\left(\overline{\Theta_1^*(\mathbf{G}_1)}^{L_\eta}\right)^\perp$. One can moreover always choose Z^f associated with $f \in F$ in the space $\{\mathbb{E}(Z|\mathcal{G}) : Z \in L_{\eta^*}\}$ of conditional expectations of r.v. in L_{η^*} given the sigma-field $\mathcal{G}$ generated by $\Theta_1^*(\mathbf{G}_1)$.

Setting $\mathbf{C} := \mathbf{C}_1 \cap \mathbf{F}$, we introduce the dual problem of (PC_y) given by

$$\textit{Maximize} \ \inf_{f\in\mathbf{C}} \langle g, f\rangle - y \int \gamma(\langle g, \theta_1\rangle)\mathrm{d}\mathbb{P} \ , \ g \in \mathbf{G}. \tag{DC$_y$}$$

The first result of this paragraph states primal attainability and primal-dual equality and it is as a simple application of part of [Léonard, 2008, Theorem 3.2]. The following functional will be useful to state and check sufficient conditions:

$$\Gamma_y^*(f) := \sup_{g\in\mathbf{G}_1} \langle g, f\rangle_{\mathbf{G}_1,\mathbf{F}_1} - y \int \gamma(\langle g, \theta_1\rangle)\mathrm{d}\mathbb{P}, \quad f \in \mathbf{F}_1. \tag{2.4.10}$$

Proposition 2.4.3 *Suppose Assumptions 2, 3 and 8 hold. Assume also that* $y > 0$ *is such that* $\mathcal{Q} \cap dom(\Psi_y)$ *is a* $\sigma(L_{\eta^*}, E_\eta)-$*weakly** *closed subset of* L_{η^*}*. Then, primal-dual equality* $(PC_y) = (DC_y)$ *holds. Moreover, if* $\mathbf{C}_1 \cap \Theta_1(dom\ (\Psi_y)) \neq \emptyset$ *or equivalently* $\mathbf{C}_1 \cap\ dom(\Gamma_y^*) \neq \emptyset$*, the minimization problem* (2.4.6) *is finite and has a unique solution* $\mathbb{Q}^y \in \mathcal{Q}$ *which satisfies*

$$\begin{aligned} \int \gamma_y^* \left(\frac{\mathrm{d}\mathbb{Q}^y}{\mathrm{d}\mathbb{P}}\right) \mathrm{d}\mathbb{P} &= \sup_{g\in\mathbf{G}} \inf_{f\in\mathbf{C}} \langle g, f\rangle - y \int \gamma(\langle g, \theta_1\rangle)\mathrm{d}\mathbb{P} \\ &= \inf_{f\in\mathbf{C}_1} \Gamma_y^*(f). \end{aligned} \tag{2.4.11}$$

Last, any minimizing sequence converges to $\mathbb{Q}^y$ *with respect to the topology* $\sigma(L_{\eta^*}, L_\eta)$*.*

Proof. The functions γ_y^* and γ_y above correspond respectively to the functions γ^* and γ in Léonard [2008] (notice that in the notation therein, we have that $m(z) = 0$ and $\gamma = \lambda$). Moreover, by Lemma 2.3.3 the functions η_y^* and η_y above correspond to the functions $\lambda_\diamond^*$ and $\lambda_\diamond$ in Léonard [2008]. Also, our mappings θ_1 and Θ_1 correspond respectively to the mappings θ and T_0 therein, and our spaces and sets $\mathbf{F}_1, \mathbf{G}_1, \mathbf{C}_1, \mathbf{F}$ and $\mathbf{G}$ correspond respectively to $\mathcal{X}_0$, $\mathcal{Y}_0$, C, $\mathcal{X}$ and $\mathcal{Y}$ in that work. One can then apply parts a) and b) of [Léonard, 2008, Theorem 3.2], conditions 1) and 2) of that result being granted by our assumptions. Notice that the equivalence between the conditions $\mathbf{C}_1 \cap \Theta_1(\text{dom }(\Psi_y)) \neq \emptyset$ and $\mathbf{C}_1 \cap \text{ dom}(\Gamma_y^*) \neq \emptyset$ follows from the "little dual equality"

$$\Gamma_y^*(f) = \inf \left\{ \Phi_y^*(Z) : Z \in L_{\eta^*}, \Theta(Z) = f \right\}, \tag{2.4.12}$$

proved in part a) of [Léonard, 2010, Proposition 5.7] (see the beginning of the proof of Theorem 2.4.2 below for an explanation of the notation used therein). ■

Remark 2.4.4 *Notice that Lemma 2.2.1 is a rewriting of identity* (2.4.12).

We will next study the attainability of the dual problem (DC_y) and characterize the measure $\mathbb{Q}^y \in \mathcal{Q}$ that solves (PC_y). We point out that, even in the reflexive setting considered here, when the convex integral function considered is not even, a solution to (DC_y) might not exist in $\mathbf{G}$, and as in Léonard [2008], a suitable extended dual problem must be considered. Attainability and characterization issues are addressed in terms of such problem in parts c) and d) of Theorem 3.2 therein. However, the fact that the function $w \mapsto \gamma_y((w)_-)$ is in null prevents us here from applying that result (which would be possible if that function and function $w \mapsto \gamma_y((w)_+)$ were both not identically null).

Nevertheless, we can adapt to our setting the study in Léonard [2008] of the extended dual problem based on abstract convex duality results of Léonard [2010]. Moreover, this study will also show that the solution to the extended dual problem actually solves (DC_y) in our case.

Let us introduce the extension of problem (DC_y), following Léonard [2008]. We denote by $\widetilde{L_\eta}$ the algebraic dual of L_{η^*}, and we write $\langle \cdot, \cdot \rangle$ for the dual product. We also consider the space $\widetilde{\mathbf{G}}$ defined as the algebraic dual of $\mathbf{F}$, and we write also $\langle \cdot, \cdot \rangle$ for the corresponding dual product (which dual product is meant should be clear from the context). Observe that the operator $\Theta_1 : L_{\eta^*} \to \mathbf{F}$ naturally induces the extension $\Theta_1^* : \widetilde{\mathbf{G}} \to \widetilde{L_\eta}$ of $\Theta_1^* : \mathbf{G} \to L_\eta$ given by

$$\langle \Theta_1^*(g), Z \rangle = \langle g, \Theta_1(Z) \rangle \,, \quad (g, Z) \in \widetilde{\mathbf{G}} \times L_{\eta^*}.$$

Introduce now the convex functions

$$\Phi_y(W) := y \int \gamma(W) \mathrm{d}\mathbb{P} \,, \quad W \in L_\eta,$$

$$\Phi_y^*(Z) := \int \gamma_y^*(Z) \mathrm{d}\mathbb{P} = \sup_{W \in L_\eta} \mathbb{E}(ZW) - y \int \gamma(W) \mathrm{d}\mathbb{P} \,, \quad Z \in L_{\eta^*}, \tag{2.4.13}$$

(the last equality by Proposition 2.3.1 a)) and

$$\overline{\Phi}_y(\zeta) := \sup_{Z \in L_{\eta^*}} \langle \zeta, Z \rangle - \Phi_y^*(Z) \,, \quad \zeta \in \widetilde{L_\eta} \,.$$

With these elements, the extended dual problem is defined as:

$$\textit{Maximize} \ \inf_{f\in\mathbf{C}}\langle g,f\rangle - \overline{\Phi}_y(\Theta_1^*(g)) \ , \ g\in\widetilde{\mathbf{G}}. \qquad (\widetilde{DC}_y)$$

We can now state the main result of this section, namely dual attainability and characterization of the primal-dual solution pairs. The function in (2.4.10) will again be useful in order to provide weak qualification conditions of purely geometric type.

Theorem 2.4.2 *Let $y>0$ and suppose all assumptions of Proposition 2.4.3 hold. Suppose moreover that $\mathbf{C}_1\cap icor\ dom(\Gamma_y^*)\neq\emptyset$.*

i) The extended dual problem $\widetilde{DC}_y$ has a solution. Moreover, any solution $\bar g$ is in $\mathbf{G}$ and thus solves (DC_y).

ii) A pair $(Z,g)\in L_{\eta^*}\times\widetilde{\mathbf{G}}$ solves PC_y and $\widetilde{DC}_y$ if and only if $g\in\mathbf{G}$ and

$$\begin{cases}\bullet\quad \Theta_1(Z)\in\mathbf{C}\cap dom\,\Gamma_y^* \\ \bullet\quad \langle g,\Theta_1(Z)\rangle\leq\langle g,f\rangle \quad \textit{for all } f\in\mathbf{C}\cap dom\,\Gamma_y^* \\ \bullet\quad Z=y\gamma'\left([\Theta_1^*(g)]_+\right).\end{cases} \qquad (2.4.14)$$

In particular, since PC_y has a unique solution and γ' is strictly increasing, any two solutions $\bar g$ and g' of $\widetilde{DC}_y$ must satisfy $[\Theta_1^*(\bar g)]_+=[\Theta_1^*(g')]_+$ a.s.

We need a preliminary result which will provide crucial information about the domain of $\overline{\Phi}_y$. It relies on [Léonard, 2008, Proposition 5.10] and on well know facts about Riesz spaces that are recalled next (we refer e.g. to [Aliprantis and Border, 2006, Chapters 8 and 9] for background).

A topological vector space L is said to be a Riesz space if it is endowed with a partial order $\leq$ making it a lattice (i.e. for any $\ell_1,\ell_s\in L$, there exists an element $\ell_1\vee\ell_2\in L$ such that $\ell_1\vee\ell_2\geq\ell_i, i=1,2$ and $\ell_1\vee\ell_2\leq\ell$ for all $\ell\in L$ such that $\ell\geq\ell_i, i=1,2$, as well as an element $\ell_1\wedge\ell_2$ defined analogously). Given $\ell\in L$, the elements ℓ_+,ℓ_- and $|\ell|$ are defined and related between each other in a similar way as in $\mathbb{R}$. The order induces an (equally denoted) dual order in the algebraic dual space of L. The space L^b of relatively bounded forms on L, also called order dual, is the subspace of the algebraic dual of L of linear forms ζ such that

$$\sup_{\ell'\in L,\,|\ell'|\leq\ell}|\langle\zeta,\ell'\rangle|<\infty \text{ for all } \ell\in L\,,\ell\geq 0.$$

By Riesz' Theorem, L^b is also a Riesz space with the dual order. In particular, elements $\zeta\in L^b$ admit (unique) positive and negative parts $\zeta_+,\zeta_-\in L^b$ such that $\zeta=\zeta_+-\zeta_-$ and $\langle\zeta_+,\ell\rangle,\langle\zeta_-,\ell\rangle\geq 0$ for all $\ell\geq 0$, with

$$\langle\zeta_+,\ell\rangle:=\sup_{\ell'\in L,\,0\leq\ell'\leq\ell}\langle\zeta,\ell'\rangle.$$

Last, recall that a normed and complete Riesz space L where its norm $\|\cdot\|$ is such that $|\ell_1| \leq |\ell_2| \Rightarrow \|\ell_1\| \leq \|\ell_2\|$ is called a Banach lattice, and that its order dual and its topological dual coincide, along with their corresponding order structures (see e.g. [Aliprantis and Border, 2006, Theorem 9.11]).

Lemma 2.4.1 *Let $\zeta \in \widetilde{L_\eta}$ be such that $\overline{\Phi}_y(\zeta) < \infty$. Then, we have $\zeta \in L_\eta$, in the sense that there exists $W^\zeta \in L_\eta$ such that $\langle \zeta, Z\rangle = \mathbb{E}(W^\zeta Z)$ for all $Z \in L_{\eta^*}$. Moreover, we have*

$$\overline{\Phi}_y(\zeta) = \overline{\Phi}_y(\zeta_+) = \Phi_y((W^\zeta)_+).$$

Proof. Let $\widetilde{L_{\eta^*}}$ denote the algebraic dual of the Riesz space L_η and $\widehat{L_{\eta^*}}$ be the corresponding subspace of relatively bounded forms. In this proof we will also denote by $\Phi_y^* : \widetilde{L_{\eta^*}} \to \mathbb{R}$ the natural extension of Φ_y^* to $\widetilde{L_{\eta^*}}$ (defined replacing the expectation in (2.4.13) by the dual product). This extension corresponds to the function Φ^* in [Léonard, 2008, Proposition 5.10], and space U therein corresponds to L_η here. We notice also that the functions Φ_+^* and Φ_-^* therein correspond respectively in our setting to the function

$$\Phi_{y,+}^*(\xi) := \sup_{W \in L_\eta} \langle \xi, W\rangle - y \int \gamma(|W|)\mathrm{d}\mathbb{P}\,, \quad \xi \in \widetilde{L_{\eta^*}},$$

and to the convex indicator of 0, since $\gamma(-|\cdot|) = 0$. It follows then from part a) of [Léonard, 2008, Proposition 5.10] that $\operatorname{dom}\Phi_y^* = \{\xi \in \widetilde{L_{\eta^*}} : \Phi_y^*(\xi) < \infty\}$ is included in $\{\xi \in \widehat{L_{\eta^*}} : \xi_- = 0\}$ and that for all $\xi \in \operatorname{dom}\Phi_y^*$, $\Phi_y^*(\xi) = \Phi_{y,+}^*(\xi_+) = \Phi_y^*(\xi_+)$.

We now show that $\operatorname{dom}\Phi_{y,+}^* \subset \widetilde{L_{\eta^*}}$ is indeed equal to L_{η^*}. The inclusion $L_{\eta^*} \subset \operatorname{dom}\Phi_{y,+}^*$ is clear since for $Z \in L_{\eta^*}$, $\Phi_{y,+}^*(Z) = \int \gamma_y^*(|Z|)\mathrm{d}\mathbb{P} < \infty$. The other inclusion is obtained by a gauge argument. More precisely, we notice that for $\xi \in \operatorname{dom}\Phi_{y,+}^*$, it holds for all non-null $W \in L_\eta$ that

$$\langle \xi, W/\|W\|_{L_\eta}\rangle \leq \Phi_{y,+}^*(\xi) + y \int \gamma(|W|/\|W\|_{L_\eta})\mathrm{d}\mathbb{P} \leq \Phi_{y,+}^*(\xi) + y,$$

since $\int \gamma(|W|/\|W\|_{L_\eta})\mathrm{d}\mathbb{P} = \int \eta(W/\|W\|_{L_\eta})\mathrm{d}\mathbb{P} \leq 1$, by definition of $\|W\|_{L_\eta}$ and Fatou's Lemma. Taking $-W$ instead of W, we then get that

$$|\langle \xi, W\rangle| \leq (\Phi_{y,+}^*(\xi) + y)\|W\|_{L_\eta},$$

which shows that $\xi \in L_{\eta^*}$.

The previous allows us to identify the spaces L and L_+ in the notation of [Léonard, 2008, Proposition 5.10] respectively defined there as the spans of $\operatorname{dom}\Phi^*$ and of $\operatorname{dom}\Phi_+^*$ with our space L_{η^*}, and the space L_- therein, defined as the span of $\operatorname{dom}\Phi_-^*$, with the trivial space $\{0\}$. Applying part b) of that result, we deduce that $\operatorname{dom}\overline{\Phi}_y = \{\zeta \in \widetilde{L_\eta} : \overline{\Phi}_y(\zeta) < \infty\}$ is included in $\widehat{L_\eta}$ and that for all $\zeta \in \operatorname{dom}\overline{\Phi}_y$ one has

$$\overline{\Phi}_y(\zeta) = \overline{\Phi}_{y,+}(\zeta_+) = \overline{\Phi}_y(\zeta_+), \tag{2.4.15}$$

where for all $\zeta \in \widetilde{L_\eta}$,

$$\overline{\Phi}_{y,+}(\zeta) := \sup_{Z \in L_{\eta^*}} \langle \zeta, Z \rangle - \Phi^*_{y,+}(Z) = \sup_{Z \in L_{\eta^*}} \langle \zeta, Z \rangle - \int \gamma^*_y(|Z|) \mathrm{d}\mathbb{P}.$$

Since the Orlicz space L_{η^*} is reflexive, by [Aliprantis and Border, 2006, Theorem 9.11] we get that $\widehat{L_\eta} = L_\eta$ so that $\operatorname{dom} \overline{\Phi}_y \subset L_\eta$ as claimed. Taking into account the fact that $\overline{\Phi}_y$ and Φ_y coincide in L_η, the remaining statements follow since the previous identification of order and topological duals is consistent with the lattice structures of these spaces. ■

Proof. (of Theorem 2.4.2)

i) Existence follows applying [Léonard, 2008, Theorem 5.3] to $\mathcal{U} = L_\eta = \mathcal{U}''$, $\mathcal{L} = L_{\eta^*}$, $\mathcal{X} = \mathbf{F}$ and $\mathcal{Y} = \mathbf{G}$ with our functions Φ_y and Θ_1 in the respective roles of functions Φ_0 and T_0 therein (notice that we have interchanged the roles of the symbols $'$ and * used to denote topological or algebraic dual spaces). The first and second properties of the solution are straightforward from Lemma 2.4.1 and the fact that the value of the primal and dual problems is real.

ii) We use [Léonard, 2008, Theorem 5.4] stating that (Z, g) is a solution if and only if the following hold:

$$\begin{cases} \bullet \quad \Theta_1(Z) \in \mathbf{C} \\ \bullet \quad \langle \Theta^*_1(g), Z \rangle_{\widetilde{L_\eta}, L_{\eta^*}} \leq \langle \Theta^*_1(g), Z' \rangle_{\widetilde{L_\eta}, L_{\eta^*}} \quad \text{for all } Z' \in \operatorname{dom} \Phi^*_y \text{ such that } \Theta_1(Z') \in \mathbf{C} \\ \bullet \quad Z \in \partial_{L_{\eta^*}} \overline{\Phi_y}(\Theta^*_1(g)). \end{cases} \tag{2.4.16}$$

Since $\gamma(-|\cdot|) = 0$, by part c) of [Léonard, 2008, Proposition 5.10] the third point is equivalent to $Z \geq 0$ and $Z \in \partial_{L_{\eta^*}} \overline{\Phi}_{y,+}(\Theta^*_1(g))$. Thus, if (Z, g) is a solution of PC_y and $\widetilde{DC}_y$ we moreover have $Z \in \partial_{L_{\eta^*}} \Phi_y([\Theta^*_1(g)]_+)$. We show now that Φ_y is Gâteaux differentiable in L_η with derivative at point $W \in L_\eta$ given by $y\gamma'(W)$. Indeed by mean value theorem and increasingness of γ we see that $\gamma(2z) - \gamma(z) \geq \gamma'(z)z$ if $z \geq 0$, and from (2.2.24) (we use that notation) we get $(k-1)\gamma(z) + d \geq \gamma'(z)z$, where necessarily $k \geq 1$. From $\gamma(W) + \gamma^* \circ \gamma'(W) = W\gamma'(W)$ we conclude that $\gamma'(W) \in L_{\eta^*}$. The claim now follows by a dominated convergence argument and recalling that the Gâteaux differential coincides with the sub-differential of a convex function when the function is differentiable in this sense. The third condition in (2.4.14) is thus satisfied. The two other conditions in (2.4.14) are straightforward using the equality (2.4.12). Reciprocally, if (2.4.14) holds, by the third point therein one has $Z \geq 0$. From there, the third point in (2.4.16) follows by similar arguments as before. ■

As a simple application of the previous result, we conclude this section with the

Proof. (of Theorem 2.2.6) We apply the previous result to the product space $\mathbf{F}_0 = \mathbb{R}^\Lambda$, and the direct sum

$$\mathbf{G}_0 = \{\alpha \in \mathbb{R}^\Lambda : \exists \Lambda' \subset \Lambda \text{ a finite subset s.t. } a_\lambda = 0 \, \forall \lambda \in \Lambda \backslash \Lambda'\},$$

the former corresponding to the algebraic dual of the latter with the product $\langle \alpha, t\rangle_{\mathbf{G}_0,\mathbf{F}_0} = \sum_{\lambda\in\Lambda'} \alpha_\lambda t_\lambda$. Moreover, we take $\theta : \Omega \to \mathbf{F}_0$ given by $\theta(\omega) = (h_\lambda(\omega))_{\lambda\in\Lambda}$, $\mathbf{C}_0 = \mathcal{C}^\Lambda$, and

$$(\nu_y : \mathbb{R}^\Lambda \times \mathbb{R} \to \mathbb{R}) = (\Gamma_y^* : \mathbf{F}_1 \to \mathbb{R}).$$

These identifications being done, part a) follows from Proposition 2.4.3. Part b) is a simple application of part a). Part c) follows immediately from part i) of Theorem 2.4.2. ■

2.5 Modular spaces and the incomplete case

In this section, the robust optimization problem without compactness in the incomplete market case will be explored. As a guideline, the Orlicz-Musielak approach is extended in an obvious way, which will lead to the so-called modular spaces. We will prove that a minimax identity, as well as existence of optimal strategies, always hold under our assumptions. On the other hand, we shall also prove that reflexivity, crucial in our approach to deriving existence of worst-case measures and characterizing the optimal strategies, is seldom obtainable, due to the presence of too many martingale measures (no matter how stringent conditions on the utility functions may be imposed).

In Subsection 2.5.1 the natural extension from the Orlicz-Musielak setting to the modular one, when one leaves the complete market case, will be motivated. Likewise the potential usefulness of this extension to the robust optimization problem will be sketched. Then in Subsection 2.5.2 and its subsequent one, the machinery of modular spaces and its link to the problem of robust optimization will be fully explored. The main result here is Theorem 2.5.1, providing a common minimax and attainability of strategies result for complete and incomplete markets, without a L^0 closedness assumption on the densities of the uncertainty set. The second crucial result is then Theorem 2.5.2 and the remarks after it, stating that reflexivity of the modular spaces under consideration extremely limits the scope of incomplete markets the theory is applicable to.

2.5.1 Modular space associated with the incomplete case

For ease of notation call $\mathcal{Y} := \{Y \in \mathcal{Y}_{\mathbb{P}}(1) : Y > 0\}$, where as usual $Y \in \mathcal{Y}$ may denote the whole process or just its end value. The Assumption 2 will always hold throughout this section.

Recall that $\eta_Y^*(z) = |z|V(Y/|z|)$ is a "random Young functions" induced by $Y \in \mathcal{Y}$. Such functions induce a space $L_{\eta_Y^*} = \{Z \in L^0 : \mathbb{E}[\eta_Y^*(\alpha Z)] < \infty, \text{ some } \alpha > 0\}$ (called, in case η_Y^* were a rho-functional, Orlicz-Musielak space) which we will denote here L_Y^* for simplicity. Again in case that η_Y^* were a rho-functional, these spaces have as previously discussed several equivalent norms, for instance the Luxemburg or the Amemiya norms respectively:

$$\|Z\|_Y^\ell := \inf\{\beta > 0 : \mathbb{E}[\eta_Y^*(\beta Z)] \le 1\} \text{ and } \|Z\|_Y^a := \inf_{k>0}\left[\frac{1}{k} + \frac{\mathbb{E}[\eta_Y^*(kZ)]}{k}\right].$$

We also define the spaces L_Y analogously, in terms of η_Y, the conjugate of η_Y^*.

Now let us define the following important functionals:

$$\begin{aligned} I(Z) &:= \inf_{Y\in\mathcal{Y}} \mathbb{E}[\eta_Y^*(Z)] &=& \inf_{Y\in\mathcal{Y}} \mathbb{E}[|Z|V(Y/|Z|)], \\ J(X) &:= \sup_{Y\in\mathcal{Y}} \mathbb{E}[\eta_Y(X)] &=& \sup_{Y\in\mathcal{Y}} \mathbb{E}[YU^{-1}(|X|)]. \end{aligned}$$

It is then clear that $v(1) = \inf_{Z\in \mathrm{d}\mathcal{Q}_e/\mathrm{d}\mathbb{P}} I(Z)$ and more generally

$$v(y) = y \inf_{Z\in \mathrm{d}\mathcal{Q}_e/\mathrm{d}\mathbb{P}} I(Z/y).$$

On the other hand, recall that the function $(Y,Z)\in (L^0)_+\times(L^0)_+ \mapsto \mathbb{E}[ZV(Y/Z)]$ is jointly convex (as $(y,z)\to zV(y/z)$ is so) and jointly lower-semicontinuous w.r.t. convergence in probability (see the proof of [Schied and Wu, 2005, Lemma 3.7]). Also recall the following Komlos-type argument (see [Delbaen and Schachermayer, 1994, Lemma A.1.1]): if $\{A_n\}_n$ is a sequence of positive random variables bounded in L^0, then there is a positive finite rv. A and a sequence $B_n\in conv\{A_n, A_{n+1},\dots\}$ such that $B_n\to A$ in probability.

We associate to the functional I a set, in complete analogy to Orlicz-Musielak spaces:

$$L_I := \left\{ Z\in L^0(\mathbb{P}) : I(\alpha Z)<\infty \text{ for some } \alpha>0 \right\}, \tag{2.5.1}$$

and define L_J accordingly in terms of J. Now we collect some elementary observations:

Lemma 2.5.1 *The following hold:*

- *The set L_I is a linear space coinciding with $\cup_{Y\in\mathcal{Y}} L_Y^*$, whereas the set L_J is a linear space contained in $\cap_{Y\in\mathcal{Y}} L_Y$.*
- *The functionals $I,J:(L^0)_+\to[0,\infty]$ are convex and moreover I is lower- semicontinuous w.r.t. convergence in measure. Also, for each non-vanishing $Z\in dom(I)$, the infimum in $I(Z)$ is attained at some $Y\in\mathcal{Y}$.*
- *$J(C)\le x \iff U^{-1}(|C|)\le X_T$ for some $X\in\mathcal{X}(x)$.*

Proof. For the convexity of I, recall that the partial infimum of every jointly convex function is convex. The fact that $I(Z)$ is attained is a consequence of the closedness and convexity of $\mathcal{Y}$, a Komlos-type argument and the lower semicontinuity of $Y\mapsto \mathbb{E}[ZV(Y/Z)]$. This in turn implies the lower semicontinuity of I, now because $(Y,Z)\mapsto \mathbb{E}[ZV(Y/Z)]$ is l.s.c.

For the first point, consider the functional I. The equality of the mentioned sets is evident from the fact that for Z fixed the infimum over the $Y\in\mathcal{Y}$ is attained. The linearity of L_I follows now from the convexity of I: if $I(\alpha Z), I(\beta X)<\infty$, taking $\gamma=\frac{\alpha\beta}{\alpha+\beta}$ yields $I(\gamma[Z+X]) = I\left[\frac{\beta}{\alpha+\beta}[\alpha Z]+\frac{\alpha}{\alpha+\beta}[\beta X]\right]\le \frac{\beta}{\alpha+\beta}I(\alpha Z)+\frac{\alpha}{\alpha+\beta}I(\beta X)<\infty$. Consider now J. That J is convex is a consequence of the convexity of U^{-1}, and from this the linearity of L_J is proved as in the case of L_I. Finally, it is clear that if $X \in L_J$ then also $X \in L_Y$, for every $Y \in \mathcal{Y}$.

The last point goes by definition of J and [Kramkov and Schachermayer, 1999, Proposition 3.1.ii]. ∎

We shall soon see that $|Z|_I^a = \inf_{k>0}[\frac{1}{k} + \frac{I(kZ)}{k}]$ is a norm on L_I, making it a Banach space. Further this norm-topology will be stronger than that of convergence in measure. This implies immediately that I would be lower-semicontinuous with respect to $|\cdot|_I^a$. In the next remark we justify the appeal of the space L_I, in light of the previous points:

Remark 2.5.1 *It has already been noted that $v(y) = y \inf_{Z \in \mathrm{d}\mathcal{Q}_e/\mathrm{d}\mathbb{P}} I(Z/y)$, but also $|Z|_I^a \leq y + yI(Z/y)$ by definition. Thus, by taking a minimizing sequence $\{Z_n\}$ such that $yI(Z_n/y)$ decreases to $v(y)$ it follows that the sequence $\{Z_n\}$ would be bounded in $(L_I, |\cdot|_I^a)$. On the other hand, we shall see in Proposition 2.5.5 that $u_{\mathbb{Q}}(x) \geq c|Z|_I^a$. This shows that in minimizing the $u_{\mathbb{Q}}$'s we may restrict $\mathcal{Q}$ to its intersection with a given ball.*

The two previous estimates show that requiring $\mathrm{d}\mathcal{Q}/\mathrm{d}\mathbb{P}$ to be closed in $(L_I, |\cdot|_I^a)$ and asking for conditions on the ingredients of the problem so that this space becomes reflexive, would allow to solve the robust optimization problem in incomplete markets along the same lines as in the complete case (we just need to replace Orlicz-Musielak spaces by the Banach space L_I). We will see, however, that L_I is reflexive almost exactly when the market is complete, and that this is independent of how well-behaved our utility function is (in stark contrast to the complete case). On the other hand, because we will be able to prove that L_J is a norm-dual space, and since the image through U of the terminal wealths live in this space, we can still obtain optimal strategies and derive a minimax identity.

2.5.2 Modular spaces L_F and E_F; topological/duality results

Generating a space from a functional is a classical subject. See e.g. Nakano [1951] and Musielak [1983]. There are quite minimalistic conditions ensuring that the generated space be an F-space and that some related functionals form a family of pseudo-norms for it. Here, rather than working at this level of generality, a more relaxed terminology and a lighter approach (as in [Nakano, 1951, Chapter XI]) will be pursued.

We first introduce the notion of a convex modular, and then its associated modular space. We shall see that I (respect. J) and L_I (respect. L_J) fulfil these definitions.

Definition 2.5.1 *A functional $F : \mathcal{S} \to [0, \infty]$ over a vector space $\mathcal{S}$ is called a Convex Modular if the following axioms are fulfilled:*

1. *$F(0) = 0$,*
2. *$F(s) = F(-s)$,*
3. *$\forall s \in \mathcal{S}, \exists \lambda > 0 : F(\lambda s) < \infty$,*

4. $F(\xi s) = 0$ *for every* $\xi > 0$ *implies* $s = 0$,

5. F *is convex,*

6. $F(s) = \sup_{0\leq\xi<1} F(\xi s)$.

Some authors would use the word modular for a much general family of functionals, and regard some of the above axioms as desirable additional properties. However with this definition, it follows that on the space

$$L_F(\mathcal{S}) := \{s \in \mathcal{S} : \lim_{\alpha\to 0} F(\alpha s) = 0\} = \{s \in \mathcal{S} : F(\alpha s) < \infty \text{ for some } \alpha > 0\},$$

the following functionals are equivalent norms, called Luxemburg and Amemiya norms respectively:

$$|s|_F^{\ell} = \inf\{\beta > 0 : F(s/\beta) \leq 1\} \quad \text{and} \quad |s|_F^{a} = \inf\left\{\frac{1}{k} + \frac{F(ks)}{k} : k > 0\right\}, \tag{2.5.2}$$

and actually thanks to [Musielak, 1983, Theorem 1.10], $|s|_F^{\ell} \leq |s|_F^{a} \leq 2|s|_F^{\ell}$. It can be proved, as in [Nakano, 1951, Chapter XI, 81], that the topology induced by the Luxemburg norm is exactly the (weakest locally convex topology) generated by the family of neighbourhoods of the origin $\{F^{-1}(-\infty, c])\}_c$. The space L_F is called a *Modular space associated to* F.

Now recalling the definitions in the previous subsection, we prove:

Proposition 2.5.1 *The functional* I *is a convex modular and* L_I *is a modular space associated to it. Likewise the functional* J *is a convex modular and* L_J *is a modular space associated to it.*

Proof. For I first. Axioms (1), (2) and (3) hold by definition, and (5) is proved in Lemma 2.5.1. For (4) notice that $I(\xi Z) = 0$ implies $\mathbb{E}[ZV(Y/(\xi Z))] = 0$ for some $Y \in \mathcal{Y}$. By positivity, this shows $ZV(Y/(\xi Z)) = 0$ a.s., from where $Z = 0$ a.s. Finally, for axiom (6), first recall that $z \mapsto zV(Y/z)$ is increasing, from which $I(Z) \geq \sup_{0\leq\xi<1} I(\xi Z) =: \zeta$. Now, take $\epsilon_n \nearrow 1$ so $\zeta = \lim I(\epsilon_n Z)$. Because I is l.s.c. we deduce that $\lim I(\epsilon_n Z) \geq I(Z)$ and thus $I(Z) = \zeta$.

Now for J. Axioms (1), (2) and (3) are direct. If $J(\xi X) = 0$ this means $YU^{-1}(\xi X) = 0$, for all $Y \in \mathcal{Y}$ a.s. Thus $X = 0$ a.s. Lastly, by increasingness of U^{-1} it holds that for fixed Y: $YU^{-1}(\xi X) \nearrow YU^{-1}(X)$ as $\xi \nearrow 1$. By monotone convergence then $\mathbb{E}[YU^{-1}(\xi X)] \nearrow \mathbb{E}[YU^{-1}(X)]$ and thus $\sup_{0\leq\xi<1} \mathbb{E}[YU^{-1}(\xi X)] = \mathbb{E}[YU^{-1}(X)]$ and now taking supremum over $Y \in \mathcal{Y}$ we get axiom (6). ■

Call now L_I^* and L_J^* the topological duals. By the "reflexivity Theorem" in Nakano [1968] it holds automatically that the modulars J and I are reflexive, in the sense that if the following functionals are defined:

$$I^*(l) := \sup_{Z\in L_I} \{l(Z) - I(Z)\} \text{ for } l \in L_I^* \text{ and } J^*(j) := \sup_{X\in L_J} \{j(X) - J(X)\} \text{ for } j \in L_J^*,$$

then I and J may be recovered, that is:

$$I(Z) = \sup_{l \in L_I^*} \{l(Z) - L^*(l)\} \text{ and } J(X) = \sup_{j \in L_J^*} \{j(X) - J^*(j)\}.$$

In particular then, both I and J are lower semicontinuous under the strong topologies introduced thus far, and by convexity, also under their weak topologies. What is more, from Lemma 2.5.2, part 1), we deduce by [Aliprantis and Border, 2006, Theorem 5.43] that both functionals are norm-continuous in the interior of their domains.

Another space of interest is the so-called set of finite elements of a modular space L_F, denoted E_F, which typically has better properties:

$$E_F = \{s \in \mathcal{S} : F(\alpha s) < \infty \text{ for all } \alpha > 0\}.$$

We remark that $E_I = L_I = dom(I)$ as soon as condition (2.2.23) in Assumption 3 holds. Let us state now a few results that will be repeatedly useful:

Lemma 2.5.2 *For every $Z \in L_I$, $X \in L_J$:*

1. $I\left(\frac{Z}{|Z|_I^\ell}\right) \leq 1$ *and* $J\left(\frac{X}{|X|_J^\ell}\right) \leq 1$
2. *Z_n norm converges to Z (respect. X_n norm converges to X) if and only if for all $\alpha > 0$, $I(\alpha[Z_n - Z]) \to 0$ (respect. $J(\alpha[X_n - X]) \to 0$)*
3. $I(Z) + J(X) \geq \mathbb{E}[XZ]$.

Proof. We prove (1) first. Notice $J\left(\frac{X}{|X|_J^\ell}\right) \leq \sup_Y \mathbb{E}[YU^{-1}(X/\|X\|_{\eta_Y}^\ell)] \leq 1$, the first inequality because clearly $|X|_J^\ell \geq \|X\|_{\eta_Y}^\ell$ and the second by definition of the Luxemburg norm and Fatou's Lemma. On the other hand take $\beta_n \searrow |Z|_I^\ell$ such that $I(Z/\beta_n) \leq 1$. Since $Z/\beta_n \to Z/|Z|_I^\ell$ in probability we conclude by Lemma 2.5.1 that $I\left(Z/|Z|_I^\ell\right) \leq \liminf I\left(Z/\beta_n\right) \leq 1$.

Part (2) is a direct consequence of [Nakano, 1951, Chapter XI.81, Theorem 3]. For part (3), by Remark 2.3.3 the conjugate of η_Y is η_Y^*, and so $\mathbb{E}[XZ] \leq \mathbb{E}[ZV(Y/Z)] + \mathbb{E}[YU^{-1}(X)]$ for every $Y \in \mathcal{Y}^*$. Thus bounding $\mathbb{E}[YU^{-1}(X)]$ above by $J(X)$ and then taking infimum over $Y \in \mathcal{Y}$ yields $\mathbb{E}[XZ] \leq I(Z) + J(X)$. ■

Time is ripe to prove some more refined properties of the spaces L_I and L_J. However, often it will be useful to lift properties satisfied by the L_Y or L_Y^* spaces to L_J and L_I respectively. Hence it is desirable that both L_Y and L_Y^* have nice properties. As it was previously discussed this is the case when they are Orlicz-Musielak spaces, which by virtue of the discussion after Lemma 2.3.1, is the case whenever $\forall\beta > 0, \mathbb{E}[V(\beta Y)] < \infty$. Call then:

$$\mathcal{Y}^* := \{Y \in \mathcal{Y} : \forall\beta > 0, \mathbb{E}[V(\beta Y)] < \infty\}.$$

To prove that there is no loss in switching from $\mathcal{Y}$ to $\mathcal{Y}^*$ it is necessary to have that $I(Z) = \inf_{Y \in \mathcal{Y}^*} \mathbb{E}[\eta_Y^*(Z)]$ and $J(X) = \sup_{Y \in \mathcal{Y}^*} \mathbb{E}[\eta_Y(X)]$. Therefore the following assumption will be made and will hold for the rest of the section unless otherwise stated:

Assumption 9 *The set $\mathcal{Y}^*$ is a non-empty subset of $\mathcal{Y}_{\mathbb{P}}(1)$ which further satisfies*

$$I(Z) = \inf_{Y\in\mathcal{Y}^*} \mathbb{E}[\eta_Y^*(Z)] \text{ , and } J(X) = \sup_{Y\in\mathcal{Y}^*} \mathbb{E}[\eta_Y(X)].$$

As said before, with this assumption L_Y and L_Y^* are Orlicz-Musielak spaces for $Y \in \mathcal{Y}^*$ and the values of I and J remain the same when considering only elements of $\mathcal{Y}^*$ in their definitions. However it should be noted that I and J may be attained by elements $Y \in \mathcal{Y}$ not necessarily in $\mathcal{Y}^*$. The next lemma gives easy sufficient conditions for the above assumption to hold.

Lemma 2.5.3 *Suppose either:*

- *that the reference measure $\mathbb{P}$ is already a martingale one*
- *that there is a continuous $\mathbb{P}-$local martingale M such that the price process is well-defined and satisfies $dS_t = dM_t + \lambda_t \cdot \mathrm{d}\langle M\rangle_t$, the market has no arbitrages and $\mathbb{E}\left[V(\beta\mathcal{E}(\lambda\cdot M)_T)\right] < \infty$ for every $\beta > 0$, where $\mathcal{E}$ stands for the stochastic exponential.*

Then Assumption 9 is verified.

Proof. For the first claim, we see easily that $1 \in \mathcal{Y}$ and naturally also $1 \in \mathcal{Y}^*$. Now take any $Y \in \mathcal{Y}$ and define $Y^n = \frac{n-1}{n}Y + \frac{1}{n}$. By convexity $Y^n \in \mathcal{Y}$, and by non-negativity $Y^n \geq \frac{1}{n}$, implying that $Y^n \in \mathcal{Y}^*$, since V is decreasing. By convexity

$$\mathbb{E}[|Z|V(Y_T^n/|Z|)] \leq \left(\frac{n-1}{n}\right)\mathbb{E}[|Z|V(Y_T/|Z|)] + \frac{1}{n}\mathbb{E}[|Z|V(1/|Z|)],$$

so $\liminf \mathbb{E}[|Z|V(Y_T^n/|Z|)] \leq \mathbb{E}[|Z|V(Y_T/|Z|)]$ and we get that $I(Z) = \inf_{Y\in\mathcal{Y}^*}\mathbb{E}[\eta_Y^*(Z)]$. On the other hand, take $X \in dom(J)$ and since of course $\frac{n-1}{n}\mathbb{E}[YU^{-1}(X)]+\frac{1}{n}\mathbb{E}[U^{-1}(X)]$ tends to $\mathbb{E}[YU^{-1}(X)]$, this directly shows that $J(X) = \sup_{Y\in\mathcal{Y}^*}\mathbb{E}[\eta_Y(X)]$. If $J(X) = +\infty$, take $\mathbb{E}[\hat{Y}_m U^{-1}(X)]$ growing to $+\infty$. If these values are finite then the previous argument shows how to approximate them in $\mathcal{Y}^*$. If (for large enough m) they are infinite, then also $\frac{n-1}{n}\hat{Y}_m + \frac{1}{n}$ generates an infinite value. Therefore the identity for J always holds.

For the second claim, and thanks to the structure of the price process, start by recalling from [Larsen and Žitković, 2007b, Proposition 3.2] that every positive process in $\mathcal{Y}$ can be written as $Y = D\mathcal{E}(\lambda\cdot M)\mathcal{E}(L)$ where L is a càdlàg local martingale strongly orthogonal to M (we write $L \in \mathcal{L}$ for this) and D is càdlàg, decreasing with $D_0 = 1$. We obviously then have that $I(Z) = \inf_{L\in\mathcal{L}}\mathbb{E}[|Z|V(\mathcal{E}(\lambda\cdot M)\mathcal{E}(L)/|Z|)]$ and $J(X) = \sup_{L\in\mathcal{L}}\mathbb{E}[\mathcal{E}(\lambda\cdot M)\mathcal{E}(L)U^{-1}(|X|))]$. Now call $\mathcal{LB} := \{L \in \mathcal{L} : \exists c > 0 \text{ st. } \mathcal{E}(L)_T \geq c\}$. We notice that $\mathcal{E}(\lambda\cdot M)\mathcal{E}(L) \in \mathcal{Y}^*$ whenever $L \in \mathcal{LB}$ since then for a $c = c(L) > 0$ we have $\mathbb{E}[V(\beta\mathcal{E}(\lambda\cdot M)_T\mathcal{E}(L)_T)] \leq \mathbb{E}[V(\beta c\mathcal{E}(\lambda\cdot M)_T)] < \infty$, by assumption. Next we follow Corollary 3.4 in Larsen and Žitković [2007b] noticing that if $L \notin \mathcal{LB}$ we may define $Y^n = \mathcal{E}(\lambda\cdot M)\left[\frac{n-1}{n}\mathcal{E}(L) + \frac{1}{n}\right]$ and get that $Y^n \in \mathcal{Y}$ from which we deduce that $Y^n = \mathcal{E}(\lambda\cdot M)\mathcal{E}(L^n)D^n$ and we easily deduce that $\mathcal{E}(L^n)_T \geq 1/n$ and thus $L^n \in \mathcal{LB}$.

Therefore:

$$\begin{aligned}\mathbb{E}\left[|Z|V\left(\frac{\mathcal{E}(\lambda\cdot M)_T\mathcal{E}(L^n)_T}{|Z|}\right)\right] &\leq \mathbb{E}\left[|Z|V\left(\frac{Y_T^n}{|Z|}\right)\right]\\ &\leq \left(\frac{n-1}{n}\right)\mathbb{E}\left[|Z|V\left(\frac{\mathcal{E}(\lambda\cdot M)_T\mathcal{E}(L)_T}{|Z|}\right)\right]\\ &+ \frac{1}{n}\mathbb{E}\left[|Z|V\left(\frac{\mathcal{E}(\lambda\cdot M)_T}{|Z|}\right)\right],\end{aligned}$$

by decreasingness and convexity of V. We thus see that

$$\liminf \mathbb{E}\left[|Z|V\left(\frac{\mathcal{E}(\lambda\cdot M)_T\mathcal{E}(L^n)_T}{|Z|}\right)\right] \leq \mathbb{E}\left[|Z|V\left(\frac{\mathcal{E}(\lambda\cdot M)_T\mathcal{E}(L)_T}{|Z|}\right)\right].$$

Hence, $I(Z) = \inf_{L\in\mathcal{LB}}\mathbb{E}[|Z|V(\mathcal{E}(\lambda\cdot M)\mathcal{E}(L)/|Z|)]$ and since $\inf_{Y\in\mathcal{Y}^*}\mathbb{E}[\eta_Y^*(Z)]$ stands in between these two values, we have the desired equality for I. Now, for J we trivially see that

$$\begin{aligned}\Delta^n &:= \mathbb{E}\left[\mathcal{E}(\lambda\cdot M)_T\mathcal{E}(L^n)_T U^{-1}(|X|)\right]\\ &\geq \left[\frac{n-1}{n}\right]\mathbb{E}\left[\mathcal{E}(\lambda\cdot M)_T\mathcal{E}(L)_T U^{-1}(|X|)\right] + \frac{1}{n}\mathbb{E}\left[\mathcal{E}(\lambda\cdot M)_T U^{-1}(|X|)\right]\\ &\to \mathbb{E}\left[\mathcal{E}(\lambda\cdot M)_T\mathcal{E}(L)_T U^{-1}(|X|)\right],\end{aligned}$$

and hence also $\mathbb{E}\left[\mathcal{E}(\lambda\cdot M)_T\mathcal{E}(L)_T U^{-1}(|X|)\right] \leq \limsup \Delta^n$.

Therefore $J(X) = \sup_{L\in\mathcal{LB}}\mathbb{E}\left[\mathcal{E}(\lambda\cdot M)_T\mathcal{E}(L)_T U^{-1}(|X|)\right]$, and again since

$$\sup_{Y\in\mathcal{Y}^*}\mathbb{E}[YU^{-1}(|X|)]$$

lies in between these two values we have the desired equality. ■

We now provide an insight into the topological properties of the spaces introduced.

Proposition 2.5.2 *Both subspaces E_I and E_J are closed subspaces of L_I and L_J respectively. When considering the almost-sure ordering, E_I and L_J are Banach lattices, and furthermore E_I is order-continuous.*

Proof. The almost-sure order is a partial order. From this both L_I and L_J are ordered vector spaces and lattices, that is, Riesz lattices. Now, because any of the norms defined in this section are lattice norm (i.e. order preserving), both L_I and L_J are Normed Riesz Spaces.

First we prove that both E_I and E_J are closed subspaces of L_I and L_J, in the spirit of the proof of [Rao and Ren, 1991, Chapter 3.4, Proposition 3]. Denote F either I or J. We need to show that $\overline{E_F} \subset E_F$. Take $s \in \overline{E_F}$ and $s_n \to s$ elements in E_F. For a fixed positive k, choose n so that $|s-s_n|_F^\ell < \frac{1}{2k}$. We then see by convexity and Lemma

2.5.2.(1), that

$$F(2k[s-s_n]) = F\left(\frac{2k[s-s_n]|2k[s-s_n]|_F^\ell}{|2k[s-s_n]|_F^\ell}\right) \leq |2k[s-s_n]|_F^\ell \leq 1.$$

Thus, since $ks = \frac{1}{2}(2k[s-s_n]) + \frac{1}{2}[2ks_n]$ we get by convexity that $F(ks) \leq \frac{1}{2}F(2k[s-s_n]) + \frac{1}{2}F(2ks_n) < \infty$. Since this holds for any $k > 0$, we conclude that $s \in E_F$. Now completeness of E_I and L_J will be proved, showing that both spaces are Banach lattices.

For E_I recall (see [Aliprantis and Border, 2006, Theorem 9.3]) that a Normed Riesz space is a Banach Lattice if and only if every positive, increasing Cauchy sequence is norm convergent. Therefore take (Z_n) a positive, increasing Cauchy sequence in E_I (for Luxemburg's norm). By definition (Z_n) converges a.s. to its supremum, which we call Z, and might be ∞-valued. Since the sequence is Cauchy, there is a $k > 0$ such that $|Z_n|_I^\ell \leq k$ for every n. By parts (1) and (3) in Lemma 2.5.2 we have that $\mathbb{E}(Z_n/k) \leq I(Z_n/k) + J(1) \leq 1 + U^{-1}(1)$ implying by Fatou's Lemma that Z is in particular finite, and so Z_n converges to Z in probability (on the non-extended real line). Notice that for every $\lambda > 0$ also $I(\lambda(Z_n - Z_m)) \to 0$ as (n,m) grows. Indeed, if $\lambda|Z_n - Z_m|_I^\ell \leq \epsilon < 1$ we have by convexity and Lemma 2.5.2.(1) that $I(\lambda(Z_n - Z_m)) \leq \lambda|Z_n - Z_m|_I^\ell \leq \epsilon$. Therefore fixing any $\lambda > 0$ we have for every $\epsilon > 0$ the existence of $N = N(\lambda, \epsilon)$ big enough s.t. $m > n > N$ implies $I(\lambda(Z_m - Z_n)) \leq \epsilon$ and hence taking limit in m by lower-semicontinuity we get $I(\lambda(Z - Z_n)) \leq \epsilon$. Thus $I(\lambda|Z_n - Z|) \to 0$ and by part (3) in Lemma 2.5.2 we see that $Z_n \to Z$ strongly. By the first part of this proof we finally get that $Z \in E_I$.

Now for L_J, take (X_n) an arbitrary Cauchy sequence. Thus the same sequence is Cauchy in every Orlicz-Musielak space associated to $YU^{-1}(\cdot)$ ($Y \in \mathcal{Y}^*$). Call $\|\cdot\|_Y$ the associated Luxemburg norm. Because these spaces are complete, the sequence norm-converges to (possibly different) limits in each of them. However, since this convergences are stronger than L^0 convergence, the limit must be necessarily (a.s.) unique. Thus, $X_n \to X$ for every Orlicz-Musielak space associated to η_Y and in probability. By Fatou's lemma $W \mapsto \mathbb{E}[YU^{-1}(W)]$ is lower-semicontinuous in $(L^0)_+$ and thus (as a supremum) also $J(\cdot)$ is so, from which $J(kX) \leq \liminf J(kX_n) \leq 1$ where k^{-1} is an upper bound for the L_J norms of the (X_n) (it exists because sequence is Cauchy) and by Lemma 2.5.2.(1). Therefore $X \in L_J$. Evidently $\|X_n - X\|_Y \leq \|X_n - X_m\|_Y + \|X_m - X\|_Y \leq |X_n - X_m|_J^\ell + \|X_m - X\|_Y$. Now given $\epsilon > 0$ we can make $|X_n - X_m|_J^\ell \leq \epsilon$ for $n, m \geq N$ independently of $Y \in \mathcal{Y}^*$. On the other hand $\|X_m - X\|_Y \leq \epsilon$ for $m \geq M(Y)$. From here, $\|X_n - X\|_Y \leq 2\epsilon$ for every $n \geq N$ independent of Y. Thus by Lemma 2.5.2.(1) again, $\mathbb{E}[YU^{-1}([X_n - X/[2\epsilon]])] \leq 1$ and taking supremum yields $J([X_n - X]/[2\epsilon]) \leq 1$ also, from which $|X_n - X|_J^\ell \leq 2\epsilon$ by definition of this norm. Therefore the sequence is convergent.

For the order-continuity of E_I, we need to show that if $Z_\alpha \searrow 0$ a.s. then $|Z_\alpha|_I \searrow 0$. Fix $\beta > 0$ and for a fixed α_0 in the set of indices, notice that $I(\beta Z_{\alpha_0}) < \infty$. Thus there is a Y such that $\mathbb{E}[Z_{\alpha_0}V(Y/(\beta Z_{\alpha_0}))] < \infty$. But $Z_\alpha V(Y/(\beta Z_\alpha))$ decreases to 0 and is dominated by $Z_{\alpha_0}V(Y/(\beta Z_{\alpha_0}))$ (for α big enough, in the sense of the net), which is integrable. By dominated (or monotone) convergence then $\mathbb{E}[Z_\alpha V(Y/(\beta Z_\alpha))] \searrow 0$ and therefore $I(\beta Z_\alpha) \searrow 0$. Since this holds for every $\beta > 0$, by Lemma 2.5.2.(3) this shows that $|Z_\alpha|_I \searrow 0$. ■

In the last result, any of the previously defined norms may have been used.

In order to further understand the modular spaces introduced thus far, and in doing so paving the way for the central statements of this section, some duality results will be pursued. First of all, Hölder-type inequalities are proved:

Proposition 2.5.3 *We have:*

$$|\mathbb{E}[XZ]| \leq |Z|_I^i |X|_J^j \leq 2|Z|_I^k |X|_J^k,$$

where $i, j, k \in \{a, \ell\}$ *and* $i \neq j$. *Furthermore, the inclusions* $L^\infty \to L_J \to L^1$ *and* $L^\infty \to L_I \to L^1$ *are continuous.*

Proof. From inequality (3) in Lemma 2.5.2 follows that $\mathbb{E}[XZ] \leq \frac{1}{\alpha\beta}\{I(\alpha Z) + J(\beta X)\}$. Now, take β such that $J(\beta X) \leq 1$. Then $\mathbb{E}[XZ] \leq \frac{1}{\beta}\left[\frac{1}{\alpha}\{1 + I(\alpha Z)\}\right]$ and taking infimum over $\alpha > 0$ yields $\mathbb{E}[XZ] \leq \frac{1}{\beta}|Z|_I^a$. Now taking infimum of the $1/\beta$ such that $J(\beta X) \leq 1$ gives $\mathbb{E}[XZ] \leq |X|_J^\ell |Z|_I^a$. From here also $|\mathbb{E}[XZ]| \leq |X|_J^\ell |Z|_I^a$ and by a similar argument $|\mathbb{E}[XZ]| \leq |X|_J^a |Z|_I^\ell$. Finally, because in the general context of Modular spaces (see [Nakano, 1951, Chapter XI]) holds that $|\cdot|^\ell \leq |\cdot|^a \leq 2|\cdot|^\ell$ we get the desired inequalities.

Evidently $1 \in L_J$ and by Assumption 9 also $1 \in L_I$. By using the derived Hölder inequalities, this shows the continuity of the inclusions into L^1. On the other hand, because both I and J are increasing, $|\cdot|_I \leq |\cdot|_\infty |1|_I$ and likewise for J, thus proving the continuity of the inclusions from L^∞. ■

Notice from this that as it can be expected, for every $X \in L_J$ the functional $l_X(\cdot) = \mathbb{E}[\cdot X]$ belongs to L_I^* and for every $Z \in L_I$ the functional $l_Z(\cdot) = \mathbb{E}[\cdot Z]$ belongs to L_J^*. Moreover, we can state a result along the lines of the Riesz representation Theorem in L^p spaces. This will rest in a few technical points to be proven in Lemma 2.5.4.

Proposition 2.5.4 *The topological dual of* E_I *is* L_J, *with the usual identification:*

$$l \in (E_I)^* \leftrightarrow l(Z) = \mathbb{E}[ZX] \text{ for some } X \in L_J,$$

and this identification is isomorphic between $(E_I, |\cdot|_I^a)$ *and* $(L_J, |\cdot|_J^\ell)$.

Furthermore, for every $Z \in L_I, X \in L_J$, *we have* $I^*(l_X) = J(X)$, *and if* $E_I = L_I$ *also* $J^*(l_Z) = I(Z)$.

Proof. Let $l \in (E_I)^*$ and define $\mu(A) := l(\mathbb{1}_A)$ for $A \in \mathcal{F}$ (well-defined and finite by Lemma 2.5.4,(1)). Clearly $\mu(\emptyset) = 0$. Also if $A_n \in \mathcal{F}$ are disjoint, and writing $A = \cup_n A_n$, then $\sum_{n \leq N} \mathbb{1}_{A_n} \to \mathbb{1}_A$ a.s. and $|\sum_{n \leq N} \mathbb{1}_{A_n} - \mathbb{1}_A| \leq 1$. Therefore by (3) in Lemma 2.5.4 then $\sum_{n \leq N} \mathbb{1}_{A_n} \to \mathbb{1}_A$ in E_I. By continuity of l then $l(\mathbb{1}_A) = \lim_N \sum_{n \leq N} l(\mathbb{1}_{A_n})$. Thus μ is clearly a finite, signed, countably-additive measure. If $A \in \mathcal{F}$ is such that $\mathbb{P}(A) = 0$ then $l(\mathbb{1}_A) = 0$ and hence $\mu(A) = 0$. So μ is absolutely continuous w.r.t. $\mathbb{P}$. By Radon-Nikodym's Theorem then $g := \frac{d\mu}{d\mathbb{P}}$ exists and is $\mathbb{P}-$integrable. By linearity then $l(f) = \mathbb{E}[fg]$ for every simple function f. By continuity $|\mathbb{E}[fg]| \leq C|f|_I$ for simple functions. Therefore $\sup\{|\mathbb{E}[fg]| : f \text{ simple and } |f|_I^a \leq 1\} < \infty$ and by (4) in Lemma 2.5.4 we get that $g \in L_J$ and that $|g|_J^\ell$ equals the above supremum. Since both $l(\cdot)$ and $\mathbb{E}(\cdot g)$ are uniformly continuous functions coinciding on a dense set (by (2) in Lemma 2.5.4, simple functions are such a set), they must agree in the whole of E_I. Hence $l(f) = \mathbb{E}[fg]$ for every $f \in E_I$ and so $(E_I)^* \subset L_J$, but using Proposition 2.5.3 the reverse inclusion already holds. Therefore $(E_I)^* = L_J$, where the identification is isomorphic if L_J is endowed with the Luxemburg norm and E_I with the Amemiya one.

Now take $X \in L_J$ and call $l_X(\cdot) := \mathbb{E}[X\cdot]$. Then:

$$\begin{aligned}
I^*(l_X) &= \sup_{Z \in L_I} \{\mathbb{E}[XZ] - \inf_{Y \in \mathcal{Y}^*} \mathbb{E}[|Z| V(Y/|Z|)]\} \\
&= \sup_{\mathcal{Y} \in \mathcal{Y}^*} \sup_{Z \in L_I} \{\mathbb{E}[XZ] - \mathbb{E}[|Z| V(Y/|Z|)]\} \\
&= \sup_{\mathcal{Y} \in \mathcal{Y}^*} \sup_{Z \in L_{\eta_Y^*}} \{\mathbb{E}[XZ] - \mathbb{E}[|Z| V(Y/|Z|)]\} \\
&= \sup_{\mathcal{Y} \in \mathcal{Y}^*} \mathbb{E}[Y U^{-1}(X)] \\
&= J(X),
\end{aligned}$$

since the conjugate of η_Y^* is η_Y. Now fix $Z \in L_I$ and assume $L_I = E_I$. Then $J^*(l_Z) = \sup_{X \in L_J}\{\mathbb{E}[XZ] - I^*[l_X]\}$ by the previous lines. On the other hand, $I(Z) = \sup_{l \in (L_I)^*}\{l(Z) - I^*(l)\} = \sup_{X \in L_J}\{\mathbb{E}[XZ] - I^*[l_X]\}$, since $(E_I)^* = L_J$. Thus $J^*(l_Z) = I(Z)$. ■

Lemma 2.5.4

1. *$\mathbb{1}_A \in E_I$ for every $A \in \mathcal{F}$*
2. *Simple functions are norm dense in E_I*
3. *If $Z_n \to 0$ a.s. and $|Z_n|$ is bounded by a constant, then $|Z_n|_I \to 0$*
4. *If $\kappa := \sup\{|\mathbb{E}[fg]| : f$ simple and $|f|_I^a \leq 1\} < \infty$ then $g \in L_J$ and $|g|_J^\ell = \kappa$*

Proof. For the first point, $\mathbb{1}_A \in E_I$ iff $\inf_{Y \in \mathcal{Y}} \mathbb{E}[\mathbb{1}_A V(\beta Y)] < \infty$ for every $\beta > 0$. This is true, simply by taking a $Y \in \mathcal{Y}^*$.

For the third point, if $|Z_n| \leq K$, then $I(\alpha Z_n) \leq \alpha \inf_{Y \in \mathcal{Y}} \mathbb{E}[K V(Y/(\alpha K))]$. But $|Z_n| V(Y/\alpha|Z_n|) \to 0$ a.s. and this sequence is dominated by $KV(Y/(\alpha K))$. Therefore if there exists a $Y \in \mathcal{Y}$ such that $\mathbb{E}[V(Y/(\alpha K))] < \infty$, then it would follow that $I(\alpha Z_n) \to 0$. But this holds (for every $\alpha > 0$) again by taking $Y \in \mathcal{Y}^*$. By Lemma 2.5.2.(3) we conclude that $Z_n \to 0$ strongly.

The proof of the second point resembles the previous one. First, since simple functions are dense in L^∞ and by Proposition 2.5.3 this last space is contained continuously in L_I (obviously then also in E_I), it suffices to show that bounded functions are dense in E_I. Take $Z \in E_I$ and define $Z_n = Z\mathbb{1}_{|Z|<n}$. Thus $X_n := |Z - Z_n| = |Z|\mathbb{1}_{|Z| \geq n} \searrow 0$ a.s. Now fix $\beta > 0$. Taking any $N > 0$ and because $\infty > I(\beta X_N) = \beta \mathbb{E}[X_N V(Y/(\beta X_N))]$ for some $Y \in \mathcal{Y}$, and $X_n V(Y/(\beta X_n)) \searrow 0$ a.s. then by dominated (or monotone) convergence $\mathbb{E}[X_n V(Y/(\beta X_n))] \to 0$ and thus $I(\beta X_n) \to 0$. Now because this holds for every β, by Lemma 2.5.2.(3) then $|X_n|_I \to 0$.

Finally, for the fourth point, take $\kappa < \infty$ as in the statement. Then clearly $\sup\{|\mathbb{E}[zg]| : z \text{ simple and } ||z||^a_{\eta^*_Y} \leq 1\} \leq \kappa$ for every $Y \in \mathcal{Y}^*$. A classical result in Orlicz theory (see (10) in [Rao and Ren, 1991, Chapter 3.4, Proposition 10]), which readily generalizes to Orlicz-Musielak spaces, implies that $\|g\|^\ell_{\eta_Y} = \sup\{|\mathbb{E}[zg]| : ||z||^a_{\eta^*_Y} \leq 1\}$, and hence $\|g\|^\ell_{\eta_Y} \leq \kappa$, since any non-negative z may be approximated in an increasing way a.s. by simple functions. Hence $\sup_{Y\in\mathcal{Y}^*} \|g\|^\ell_{\eta_Y} \leq \kappa$. Since $\mathbb{E}[YU^{-1}(g/\|g\|^\ell_{\eta_Y})] \leq 1$ (by definition of the norm and Fatou's Lemma) then $\mathbb{E}[YU^{-1}(g/\kappa)] \leq 1$ and thus $J(g/\kappa) \leq 1$ from which $|g|^\ell_J \leq \kappa < \infty$. Finally, by Proposition 2.5.3 we have $|\mathbb{E}[fg]| \leq |g|^\ell_J |f|^a_I$ and so if f is simple and such that $|f|^a_I \leq 1$ we get that $|\mathbb{E}[fg]| \leq |g|^\ell_J$ and then by taking supremum over such functions we derive that $\kappa \leq |g|^\ell_J$, and therefore there is equality.
■

Notice that being $L_J = (E_I)^*$ this already implies completeness of L_J. Also notice that a property analogous to point (3) in the above lemma does not hold in E_J if $\mathcal{Y}$ is not uniformly integrable. This already is evidence of non-reflexivity in a general situation. We shall discuss this in depth later on.

2.5.3 Applications of the Modular approach to the robust optimization problem

We are ready to apply the previously introduced modular spaces in the study of the robust optimization problem. In what follows it will be assumed that the uncertainty set $\mathcal{Q}$ satisfies the following assumptions:

Assumption 10

- $\mathcal{Q}$ *is countably convex.*
- $[\mathbb{P}(A) = 0 \iff \forall \mathbb{Q} \in \mathcal{Q}, \mathbb{Q}(A) = 0]$.
- $\mathrm{d}\mathcal{Q}/\mathrm{d}\mathbb{P}$ *is a non-empty $\sigma(L_I, L_J)$-closed subset of L_I.*
- $\exists x > 0, \mathbb{Q} \in \mathcal{Q}_e$ *such that $u_\mathbb{Q}(x) < \infty$.*

As commented before, the first two assumptions imply the non-emptiness of the set $\mathcal{Q}_e$ of equivalent measures in $\mathcal{Q}$. We could actually replace, as will be mentioned in Remark 2.5.2, the first and fourth points by [$\mathcal{Q}$ is convex and $\mathrm{d}\mathcal{Q}_e/\mathrm{d}\mathbb{P} \cap L_I \neq \emptyset$].

As a consequence of Proposition 2.5.3, we can prove the following result, of interest on its own, which we already mentioned in Remark 2.5.1 and will be useful in proving the general minimax Theorem 2.5.1 below:

Proposition 2.5.5 *Suppose Assumptions 2 and 10. Then, for all $x > 0$ we have that*

$$\forall \mathbb{Q} \in \mathcal{Q}: \quad (1+x)\left|\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right|^\ell_I \geq u_\mathbb{Q}(x) \geq (1\wedge x)\left|\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right|^a_I. \tag{2.5.3}$$

Proof. By Proposition 2.5.3 we have:

$$\mathbb{E}^{\mathbb{Q}}[U(X_T)] \leq |\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}|_I^{\ell} |U(X_T)|_J^{a} \leq [1 + J(U(X_T))] |\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}|_I^{\ell},$$

by definition of the norm. Hence, by Lemma 2.5.1 we get that $u_{\mathbb{Q}}(x) \leq [1+x]|\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}|_I^{\ell}$. Now we prove the lower bound for $u_{\mathbb{Q}}(x)$ in (2.5.3). Let us call $Z = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} \in \frac{\mathrm{d}\mathcal{Q}_e}{\mathrm{d}\mathbb{P}}$. Recalling that $v_{\mathbb{Q}}(y) := \inf_{Y \in \mathcal{Y}} \mathbb{E}\left[ZV(yY/Z)\right]$, we have:

$$|Z|_I^a \leq y + yI(Z/y) = y + v_{\mathbb{Q}}(y) \leq y + cv_{\mathbb{Q}}(y),$$

for each $c \geq 1$. Calling $A_{\mathbb{Q}}(y) = v_{\mathbb{Q}}(y) + xy$, then $A_{\mathbb{Q}}(y) \geq \frac{1}{c}|Z|_I + \left(x - \frac{1}{c}\right) y$. Thus for every $x > 0$, finding $c \geq 1$ such that $x \geq c^{-1}$ and then taking infimum over $\{y > 0\}$ yields $u_{\mathbb{Q}}(x) \geq C|Z|_I$: if the r.h.s. is infinite there is nothing to prove, and otherwise by [Kramkov and Schachermayer, 1999, Theorem 3.1] it holds $u_{\mathbb{Q}}(x) = \inf_{y>0}\left[v_{\mathbb{Q}}(y) + xy\right]$ and we still get the desired bound. The best constant C is thus $1 \wedge x$.

If now $Z := \mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P} \in \frac{\mathrm{d}\mathcal{Q}}{\mathrm{d}\mathbb{P}} \setminus \frac{\mathrm{d}\mathcal{Q}_e}{\mathrm{d}\mathbb{P}}$ is such that $u_{\mathbb{Q}}(x) = +\infty$, the lower bound trivially holds. If in turn $u_{\mathbb{Q}}(x) < \infty$ we resort to [Schied and Wu, 2005, Lemma 3.3], which states that whenever $\mathbb{Q}_i \in \mathcal{Q}, i = 1, 2$ are such that $u_{\mathbb{Q}_i} < \infty$, the function $t \in [0,1] \mapsto u_{t\mathbb{Q}_1 + (1-t)\mathbb{Q}_2}(x)$ is continuous $\forall x > 0$ (note that this result does not use the assumption that $\frac{\mathrm{d}\mathcal{Q}}{\mathrm{d}\mathbb{P}}$ be closed in L^0). Notice that $Z_t = t\mathrm{d}\mathbb{Q}_0/\mathrm{d}\mathbb{P} + (1-t)Z$ (with $\mathbb{Q}_0$ as in the statement of the present result) satisfies $Z_t \in \frac{\mathrm{d}\mathcal{Q}_e}{\mathrm{d}\mathbb{P}}$ for $0 < t \leq 1$. Thus we get that $\forall \epsilon > 0, \exists \delta$ such that $t \in (0, \delta) \Rightarrow u_{\mathbb{Q}}(x) \geq u_{Z_t \mathrm{d}\mathbb{P}}(x) - \epsilon \geq C_x |Z_t|_I^a - \epsilon$. Hence, taking $t \to 0+$ we get that $u_{\mathbb{Q}}(x) \geq C_x |Z|_I - \epsilon, \forall \epsilon > 0$, and the lower bound follows by letting $\epsilon \to 0$. ∎

Remark 2.5.2 *In light of the upper bound in* (2.5.3)*, we see that indeed the third and fourth points in Assumption 10 could be replaced respectively by* $\mathrm{d}\mathcal{Q}/\mathrm{d}\mathbb{P} \cap L_I$ *being weakly closed and* $\mathrm{d}\mathcal{Q}_e/\mathrm{d}\mathbb{P} \cap L_I \neq \emptyset$*, and so likewise the first point by the usual convexity of the set* $\mathcal{Q}$*.*

Thanks to Proposition 2.5.4 we can endow L_J with a weak-* topology and thus prove the following crucial minimax Theorem, of which Theorems 2.2.3 and 2.4.1 are special cases:

Theorem 2.5.1 *Suppose Assumptions 2, 9 and 10 hold. Assume that* $L_I^* \cong L_J$ *(e.g. under condition* (2.2.23) *in Assumption 3). Then for every* $x > 0$*:*

$$\begin{aligned} u(x) \quad &= \inf_{\mathbb{Q} \in \mathcal{Q}} \sup_{X \in \mathcal{X}(x)} \mathbb{E}^{\mathbb{Q}}\left(U\left(X_T\right)\right) \quad = \inf_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}^{\mathbb{Q}}\left(U\left(\hat{X}_T\right)\right) \\ &= \inf_{\mathbb{Q} \in \mathcal{Q}_e} \sup_{X \in \mathcal{X}(x)} \mathbb{E}^{\mathbb{Q}}\left(U\left(X_T\right)\right) \quad < +\infty, \end{aligned} \tag{2.5.4}$$

for some $\hat{X} \in \mathcal{X}(x)$*, and moreover* v *is finite and* u, v *are conjugate on* $(0, \infty)$*.*

Furthermore, if L_I *is reflexive (e.g. in the complete market case under the two conditions in Assumption 3) there is a saddle point, i.e. there exists a* $\hat{\mathbb{Q}} \in \mathcal{Q}$ *so that all the above values equal* $\mathbb{E}^{\hat{\mathbb{Q}}}\left[U\left(\hat{X}_T\right)\right]$*.*

Proof. Fix $x > 0$. We now intend to apply [Aubin and Ekeland, 2006, Chapter 6, Theorem 7] (Lopsided minimax Theorem, also stated on page 295 therein). First, let us define the set $G := \{g \in L_J : 0 \leq g \leq U(X_T), \text{ some } X \in \mathcal{X}(x)\}$. Now let us define a bilinear function $F : G \times \mathrm{d}\mathcal{Q}/\mathrm{d}\mathbb{P} \to [0, \infty)$ by $F(g, Z) = \mathbb{E}[Zg]$. Evidently under condition $L_I^* \cong L_J$ we must have that $E_I = L_I$ (which is the case anyway if condition (2.2.23) in Assumption 3 holds).

We first endow the convex set G with the weak-* topology $\sigma(L_J, E_I)$. Let us prove that G is closed with it. Indeed if $\{g_\alpha\}_\alpha \subset G$, we have by Lemma 2.5.1, part c), that $J(g_\alpha) \leq x$. But by Proposition 2.5.4, the spaces $(E_I, \sigma(E_I, L_J)), (L_J, \sigma(L_J, E_I))$ are in topological duality and $J = I^*$. Therefore J is $\sigma(L_J, E_I)$-l.s.c. and we conclude that if $g_\alpha \to g$ in this topology, then $J(g) \leq x$. Again by Lemma 2.5.1, part c), we see that $|g| \in G$. On the other hand $\mathbb{1}_{g<0} \in E_I$ (by Lemma 2.5.4) and so $\mathbb{E}[g\mathbb{1}_{g<0}] = \lim \mathbb{E}[g_\alpha \mathbb{1}_{g<0}] \geq 0$, from which $g \geq 0$ and so $g \in G$.

We now prove that G is weak*-compact. By Banach-Alaoglu it suffices to prove that it is norm bounded. But this holds since $|g|_J^a \leq 1 + J(g) \leq 1 + x$, for every $g \in G$.

We apply now the lopsided minimax Theorem. The function F satisfies:

- $F(g, \cdot)$ is convex
- $\{g \in G : F(g, Z) \geq \beta\}$ is weak*-compact for every β, Z.
- $F(\cdot, Z)$ is concave and continuous,

and thus $-F$ satisfies with ease the requirements of that theorem. We conclude then the minimax equality and the attainability of an optimal $g \in G$. By simple arguments in Schied and Wu [2005] (see the proof of Lemma 3.4 therein) any optimal g must be of the form $U(X_T)$ and one may approximate the "infsup" by taking the infimum over $\mathcal{Q}_e$.

Because we proved that $u(x) = \inf\limits_{\mathbb{Q} \in \mathcal{Q}_e} u_\mathbb{Q}(x)$ we also have $u(x) = \inf\limits_{\mathbb{Q} \in \mathcal{Q}_e, u_\mathbb{Q}(x) < \infty} u_\mathbb{Q}(x)$. Now applying [Kramkov and Schachermayer, 1999, Theorem 3.1] we see that $u(x) = \inf_{y \geq 0} \left[\inf_{\mathbb{Q} \in \mathcal{Q}_e, u_\mathbb{Q}(x) < \infty} v_\mathbb{Q}(y) + xy\right]$ and so by the first statement in [Schied and Wu, 2005, Lemma 3.5] we conclude that u is the conjugate of v. Finiteness of v on $(0, \infty)$ is a consequence of $L_I = E_I$. Because I is convex and $v(y) = \inf_{Z \in \mathrm{d}\mathcal{Q}_e/\mathrm{d}\mathbb{P}} yI(Z/y)$, an argument as in the proof of Lemma 2.3.2 shows that v is convex and so we conclude by [Aliprantis and Border, 2006, Theorem 7.22] that v is continuous in $(0, \infty)$. Since clearly $v(y) \geq V(y)$ we see that $v(0+) = \infty$. Thus defining $v(\cdot) = \infty$ on $(-\infty, 0]$ we get a l.s.c. function everywhere. Defining $u(0) = 0$ and $u(x) = -\infty$ if $x < 0$, we still get that u is the concave conjugate of v. This in turn implies that v is conjugate to u and also that if $y > 0$ then $v(y) = \sup_{x > 0}[u(x) + xy]$.

Finally, in the reflexive case, when computing $\inf_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}^\mathbb{Q}\left(U\left(\hat{X}_T\right)\right)$ we realize that it is enough to do it over a norm-bounded subset of $\mathrm{d}\mathcal{Q}/\mathrm{d}\mathbb{P}$. Indeed, we have already proven that $u(x) = \inf_{\mathbb{Q} \in \mathcal{Q}} u_\mathbb{Q}(x)$, and this is finite by Assumption 10. Thus we may only regard $\mathcal{Q} \cap \{\mathbb{Q} : u_\mathbb{Q}(x) \leq u(x) + 1\}$, but by Proposition 2.5.5 we have that $u_\mathbb{Q}(x) \geq c(x)|\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}|_I^a$, and so this set is contained in $\mathcal{Q} \cap \{\mathbb{Q} : |\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}|_I^a \leq c(x)^{-1}[u(x) + 1]\}$. By reflexivity and Assumption 10, these sets are weakly compact (i.e. $\sigma(E_I, L_J)$-compact) and so the continuous linear functional $Z \mapsto \mathbb{E}\left(ZU\left(\hat{X}_T\right)\right)$ attains its minimum there. Any of these densities along with any optimal $\hat{X}$ conforms a saddle point.

We finally stress that the reflexivity condition on L_I is satisfied if the market is complete and Assumption 3 holds. Indeed by completeness we would have that $I(\cdot) = \mathbb{E}[\eta_1^*(\cdot)]$ and $J(\cdot) = \mathbb{E}[\eta_1(\cdot)]$, and so by Assumption 3 coupled with Proposition 2.3.1 and Theorem 2.3.2 we get the desired reflexivity. ■

Remark 2.5.3 *Proposition 2.5.5 and the previous theorem prove that under the assumption $\exists x > 0, \mathbb{Q} \in \mathcal{Q}_e$ such that $u_{\mathbb{Q}}(x) < \infty$, it is equivalent to write closedness conditions on $\mathrm{d}\mathcal{Q}/\mathrm{d}\mathbb{P} \cap L_I$ or on $\mathrm{d}\mathcal{Q}/\mathrm{d}\mathbb{P} \subset L_I$.*

Remark 2.5.4 *From the previous proof it is clear that if $\mathrm{d}\mathcal{Q}/\mathrm{d}\mathbb{P} \subset E_I$ then at least for the minimax result and the existence of an optimal wealth, the condition $L_I^* \cong L_J$ can be avoided altogether, since we may work on E_I instead if L_I from the beginning.*

Let us point out that at the moment we can only prove existence of a worst-case $\hat{\mathbb{Q}}$ (as well as relating it explicitly to the optimal $\hat{X}$) in the case that our modular spaces are reflexive. In Theorem 2.5.2 and Remark 2.5.6, we aim to find out when this is the case. The following property relates the answer to the set $\mathcal{Y}$.

Lemma 2.5.5 *If E_J has order-continuous norm (i.e. $|x_\alpha|_J \searrow 0$ whenever $x_\alpha \searrow 0$) then $\mathcal{Y}$ is uniformly integrable.*

Proof. By [Aliprantis and Border, 2006, Theorem 9.22], E_J has order-continuous norm if and only if every sequence of order-bounded and disjoint elements is strongly convergent to zero. So take A_n a sequence of disjoint sets. Notice that $\mathbb{1}_{A_n}$ is an order-bounded and disjoint sequence, and thus $|\mathbb{1}_{A_n}|_J \to 0$. This implies $J(\mathbb{1}_{A_n}) \to 0$, which means $\sup_{Y \in \mathcal{Y}} \mathbb{E}[\mathbb{1}_{A_n} Y] \to 0$. Now, from [Diestel, 1991, Theorem 7] this implies that $\mathcal{Y}$ is uniformly integrable. ■
The following theorem is essential and it implies Theorem 2.2.4.

Theorem 2.5.2 *If the set $\mathcal{Y}$ is not uniformly integrable, then neither E_J, L_J nor E_I can be reflexive.*

Proof. As pointed out in [Aliprantis and Border, 2006, Corollary 9.23], a reflexive Banach lattice has order continuous norm. Since E_J is a Banach lattice in itself, if it were reflexive, by Lemma 2.5.5 the set $\mathcal{Y}$ would be uniformly integrable. Thus E_J is not reflexive and therefore L_J neither, since the former is a closed subset of the latter. On the other hand, under the assumption of this section the dual of E_I is isomorphic to L_J (which we proved in Proposition 2.5.4) which in turn implies that E_I cannot be reflexive either. ■

Remark 2.5.5 *The previous result states that lack of uniform integrability of $\mathcal{Y}$ implies that the space L_I cannot be reflexive. This lack of reflexivity means that the approach used for Orlicz-Musielak spaces (in the complete case) does not extend vis-à-vis to the* *current Modular space setting, since subsequence principles rely precisely on reflexivity (see Remark 2.5.1 for some context). It is remarkable that no growth conditions on U or V may yield reflexivity to our Modular spaces as soon as $\mathcal{Y}$ is not uniformly integrable.*

Remark 2.5.6 *If the set $\mathcal{Y}$ were uniformly integrable, then also the set of absolutely continuous martingale measure $\mathcal{M}$ would be so (more precisely, their densities would be $\sigma(L^1, L^\infty)-$ relatively compact). The results [Delbaen, 1992, Theorem 6.7 and Corollary 7.2] then say that $\mathcal{M}$ must be a singleton, at least in the case of bounded continuous prices and either if all martingales on the filtration are continuous (this is the case of the augmented brownian filtration) or if the filtration is quasi left-continuous. Therefore in most interesting cases uniform integrability of $\mathcal{Y}$ implies market completeness.*

We have seen in Theorem 2.5.1 how despite the lack of reflexivity the modular space approach still provides a nice setting to tackle robust problems without a-priori compactness. We close this section remarking that this framework may also be useful in other contexts (e.g. risk measures, pricing, etc). To illustrate the point, we derive a rather short proof of the existence of an optimal strategy in the classical, non-robust, utility maximization problem in incomplete markets (compare with [Kramkov and Schachermayer, 1999, Lemma 3.9], [Kramkov and Schachermayer, 2003, Lemma 1], or the *proof of existence* after [Schachermayer, 2004, Lemma 3.16]) by means of the modular spaces and weak compactness. This can be seen of course as corollary to our minimax Theorem, but for the sake of the argument we show the full proof:

Proposition 2.5.6 *Under Assumptions 2 and 9, the value function $u_{\mathbb{P}}(\cdot)$ is everywhere finite and for every $x > 0$ the problem $u_{\mathbb{P}}(x) := \sup_{X \in \mathcal{X}(x)} \mathbb{E}[U(X_T)]$ is attained at a unique element.*

Proof. Since $\mathcal{Y}^* \neq \emptyset$ we get that L_I contains the constants and thus by the upper bound in Proposition 2.5.5 we get that $u_{\mathbb{P}}(\cdot) < \infty$.

Fix now $x > 0$ and take $X_n \in \mathcal{X}(x)$ so that $\mathbb{E}[U(X^n_T)] \to u_{\mathbb{P}}(x)$. By definition we derive that $|U(X^n_T)|^a_J \leq 1 + J(U(X^n_T)) \leq 1 + x$, where we also used Lemma 2.5.1. By Proposition 2.5.4 and Banach-Alaoglu Theorem, there is a subnet of $U(X_n)$ convergent in $\sigma(L_J, E_I)$ (i.e. weak*) to an element $K \in L_J$. Since $1 \in E_I$ by Assumption 9 we conclude that $u(x) = \mathbb{E}[K]$. Further, by invoking part (1) of Lemma 2.5.4 we conclude that $\mathbb{E}[K \mathbb{1}_{K<0}] \geq 0$ and so $K \geq 0$ a.s. On the other hand, Proposition 2.5.4 also shows that $(E_I, \sigma(E_I, L_J))$ and $(L_J, \sigma(L_J, E_I))$ are in separating duality and have compatible topologies, and by the same Proposition $J = (\tilde{I})^*$ where $\tilde{I}$ is I's restriction to E_I. This proves that J is $\sigma(L_J, E_I)$-l.s.c. and in particular $J(K) \leq x$, since $J(U(X^n_T)) \leq x$ (this from Lemma 2.5.1). By the same lemma we deduce that $U^{-1}(K) \leq \hat{X}_T$ for some $\hat{X} \in \mathcal{X}(x)$ and so $K \leq U(\hat{X}_T)$, which proves that $u_{\mathbb{P}}(x) = E[U(\hat{X}_T)]$.

Uniqueness is a consequence of U being strictly concave. ■

Remark 2.5.7 *The previous result may appear to contradict the statement in Kramkov and Schachermayer [1999] that the condition $AE(U) < 1$ is almost necessary for the existence of optimal wealth, but this is not the case. Concretely, in Example 5.2 therein the* *authors construct a complete market for which whenever* $AE(U) = 1$ *no optimal wealth processes exist for initial capital* x *big enough. However in this type of constructions one needs that* $\mathbb{E}[V(\beta Y)] = +\infty$ *for every* β *small enough (see Lemma 5.1.iii therein), where* Y *is the unique martingale measure, and so violate the condition* $\mathcal{Y}^* \neq \emptyset$ *in our Assumption 9. Thus our assumptions rule out the combinations of market models and utility functions for which such counter-examples can be built. This is in accordance with Kramkov and Schachermayer [2003] where sufficient and necessary conditions are given for pairs of market models and utility functions.*

Remark 2.5.8 *We envisage that further analysis of our modular spaces (for instance identifying the dual of* L_J*) may bring more understanding of the robust problem and the (non)existence of the associated worst-case measures. This could be endeavoured through minimization of entropy techniques as well. We should also say that one may add more ad-hoc conditions on the set* $\mathrm{d}\mathcal{Q}/\mathrm{d}\mathbb{P}$ *in order to retrieve in incomplete markets the results obtained for the complete one, even without reflexivity. An example would be to ask* $\mathrm{d}\mathcal{Q}/\mathrm{d}\mathbb{P} \cap B_{L_I}(0, R)$ *to be weakly compact, for* R *large enough. Then most of the points in Theorems 2.2.2 and 2.2.1 could be derived. However, tractable conditions implying such a property should still be provided in order to derive applicable criteria.*

2.6 On a possible extension to the non-dominated case

In this last brief section of the present chapter we will discuss heuristically how its core idea, namely that we can relax the requirements on the uncertainty set thanks to a certain compactness on the set of images through the utility function of the admissible wealths, may be generalized for the case of non-dominated models. This should serve as a motivation for that line of research, and also highlight where the hardest difficulties should arise when implementing the suggested programme.

We start by remarking that in the previous sections there was always a privileged probability measure, the reference one $\mathbb{P}$, and that all the measures in the uncertainty set were assumed to be absolutely continuous with respect to it. In fact, if we define the set function $\mathcal{C}_{\mathcal{Q}}(\cdot) = \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{Q}(\cdot)$, we as well as Schied and Wu [2005] had assumed that "$\mathcal{C}_{\mathcal{Q}}$ is equivalent to $\mathbb{P}$," in the sense that $\mathcal{C}_{\mathcal{Q}}(A) = 0 \iff \mathbb{P}(A) = 0$.

This is however not an appropriate assumption for models with misspecified (i.e. uncertain) volatility. For instance, to exemplify, it is well-known that the measures on path-space making the coordinate process a martingale with density of its quadratic variation w.r.t. the Lebesgue measure equal to infinite different constants are actually singular to each other and non-dominated. Thus, the case when the uncertainty set $\mathcal{Q}$ is not dominated by a single measure is of great interest, and indeed there has been a lot of work in the subject lately. We will often refer to Denis and Kervarec [2013] and do not provide here an extensive list of the research that has been done on the issue.

In this setting, which we refer to as the *non-dominated case*, already talking about stochastic integrals is a delicate subject. We avoid discussing these technicalities and refer to Soner et al. [2011], Denis and Martini [2006], Denis and Kervarec [2013], Nutz

[2012] for some possible ways to give a meaningful sense of a stochastic integral well-defined for every $\mathbb{Q} \in \mathcal{Q}$ simultaneously. In this section we keep the notation we have used before and as usual Assumption 2 is supposed to hold. We will also use the set function $\mathcal{C}_{\mathcal{Q}}$ but do not assume it absolutely continuous w.r.t. any measure. As a matter of language, a set A such that $\mathcal{C}_{\mathcal{Q}}(A) = 0$ will be called polar. We remark that we do not assume a priori that the set $\mathcal{Q}$ is weakly-compact (i.e. tight), as opposed to Denis and Kervarec [2013].

Let us assume then that we can define the sets of non-negative admissible terminal wealths $\mathcal{X}(x)$, in a filtered space $(\Omega, \{\mathcal{F}_t\}_{0\leq t\leq T})$, so that the expressions $\mathbb{E}^{\mathbb{Q}}[U(X_T)]$ are meaningful. Further let us denote by $\mathcal{P}_m$ the set of probability measures on $\mathcal{F}_T$ under which every element in $\mathcal{X} := \bigcup_x \mathcal{X}(x)$ is a martingale. As before we are interested in the problem

$$u(x) := \sup_{X\in\mathcal{X}(x)} \inf_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}}[U(X_T)],$$

the difference being that now we cannot write $\mathbb{E}^{\mathbb{Q}}[U(X_T)] = \mathbb{E}^{\mathbb{P}}[U(X_T)\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}]$ because there is no reference $\mathbb{P}$. So, if we want to do anything along the lines of what was successful in the dominated case, we need to identify first of all the modular functionals.

For a signed measure of finite variation $\mathbb{Q}$ we denote by $|\mathbb{Q}|$ its associated variation measure and $\mathcal{P}(\mathbb{Q})$ the subset of $\mathcal{P}_m$ consisting of elements absolutely continuous w.r.t. $|\mathbb{Q}|$ and we start by looking at the trivial inequality:

$$u(x) \leq \inf_{y>0}\left(\inf_{\mathbb{Q}\in\mathcal{Q}}\ \inf_{Z\in\mathrm{d}\mathcal{P}(\mathbb{Q})/\mathrm{d}|\mathbb{Q}|} \mathbb{E}^{|\mathbb{Q}|}\left[V\left(yZ\right)\right] + xy\right),$$

where for $\mathbb{Q} \in \mathcal{Q}$ obviously $\mathbb{Q} = |\mathbb{Q}|$, but we use the latter to motivate the definition of I below. We notice that the r.h.s. suggests that we should look at a functional of $\mathbb{Q}$ that already summarizes the presence of the martingale densities.

We believe that in the present case something along the lines of

$$I(\mathbb{Q}) \ := \begin{cases} 0 & \text{if } \mathbb{Q} = 0 \\ +\infty & \text{if } \mathcal{P}(\mathbb{Q}) = \emptyset \\ \inf\left\{\mathbb{E}^{|\mathbb{Q}|}\left[V(Z)\right] \ : \ Z \in \overline{\frac{\mathrm{d}\mathcal{P}(\mathbb{Q})}{\mathrm{d}|\mathbb{Q}|}}^{L^0(|\mathbb{Q}|)}\right\} & \text{otherwise,} \end{cases}$$

is likely appropriate. As explained in [Schachermayer, 2004, p.48], we could take instead of $\overline{\frac{\mathrm{d}\mathcal{P}(\mathbb{Q})}{\mathrm{d}|\mathbb{Q}|}}^{L^0(|\mathbb{Q}|)}$ the sets

$$\left\{Z : 0 \leq Z \leq W,\ \mathbb{Q}\text{-a.s., some } W \in \overline{\frac{\mathrm{d}\mathcal{P}(\mathbb{Q})}{\mathrm{d}|\mathbb{Q}|}}^{L^0(|\mathbb{Q}|)}\right\}, \text{ or}$$

$$\mathcal{Y}_{\mathbb{Q}}(1) := \{Y \text{ non-negative } |\mathbb{Q}|\text{-supermart.} : Y_0 = 1, XY \text{ is a } |\mathbb{Q}|\text{-supermart.}, \forall X \in \mathcal{X}\},$$

where as usual $\mathcal{Y}_{\mathbb{Q}}(1)$ may also denote the set of final values of the processes therein (context should be clear). The functional I is defined in the space of all signed measures of finite variation, and here and elsewhere $L^0(|\mathbb{Q}|)$ denotes the space of measurable functions (identifying $|\mathbb{Q}|$-a.e. equal ones) equipped with convergence in $|\mathbb{Q}|$-measure, and we keep the notation $\mathbb{E}^{|\mathbb{Q}|}$ even-though the measure might not have mass equal to one. It is easily seen that in case we had a reference measure $\mathbb{P}$, the previous functional coincides with the one we had in the dominated case (called I likewise) on the subset of signed measures of finite variation given by $\{Z\mathrm{d}\mathbb{P} : Z \in L^1(\mathbb{P})\}$. Furthermore, upon defining $v(y) = \inf_{\mathbb{Q}\in\mathcal{Q}} \inf_{Y\in\mathcal{Y}_{\mathbb{Q}}(y)} \mathbb{E}^{\mathbb{Q}}\left(V\left(Y_T\right)\right)$, it is then clear that $v(1) = \inf_{\mathbb{Q}\in\mathcal{Q}} I(\mathbb{Q})$ and more generally $v(y) = y \inf_{\mathbb{Q}\in\mathcal{Q}} I(\mathbb{Q}/y)$.

Let us now consider the set of signed measures of finite variation absolutely continuous w.r.t. $\mathcal{C}_{\mathcal{Q}}$, denoted $\mathcal{M}(\mathcal{C}_{\mathcal{Q}}) := \{\mathbb{Q} \text{ measure } : |\mathbb{Q}|(\Omega) < \infty, \mathbb{Q} \ll \mathcal{C}_{\mathcal{Q}}\}$, where $\mathbb{Q} \ll \mathcal{C}_{\mathcal{Q}}$ means $|\mathbb{Q}| \ll \mathcal{C}_{\mathcal{Q}}$ (which we already defined). We restrict preliminarily

$$I : \mathcal{M}(\mathcal{C}_{\mathcal{Q}}) \to \mathbb{R}_+ \cup \{+\infty\},$$

and define the sets:

$$\begin{aligned} L_I &:= \{\mathbb{Q} \in \mathcal{M}(\mathcal{C}_{\mathcal{Q}}) : \exists \alpha > 0, I(\alpha\mathbb{Q}) < \infty\}, \\ E_I &:= \{\mathbb{Q} \in \mathcal{M}(\mathcal{C}_{\mathcal{Q}}) : \forall \alpha > 0, I(\alpha\mathbb{Q}) < \infty\}. \end{aligned}$$

By inspection we also define $J : \mathcal{L}(\mathcal{C}_{\mathcal{Q}}) \to \mathbb{R}_+ \cup \{+\infty\}$ by

$$J(X) := \sup_{\mathbb{Q}\in L_I, Y\in\mathcal{Y}_{\mathbb{Q}}} \mathbb{E}^{|\mathbb{Q}|}\left[YU^{-1}(|X|)\right],$$

where $\mathcal{L}(\mathcal{C}_{\mathcal{Q}})$ denotes the space of equivalence classes of measurable functions where polar equal ones (w.r.t. $\mathcal{C}_{\mathcal{Q}}$) are identified. Let us introduce accordingly:

$$\begin{aligned} L_J &:= \{X \in \mathcal{L}(\mathcal{C}_{\mathcal{Q}}) : \exists \alpha > 0, J(\alpha X) < \infty\}, \\ E_J &:= \{X \in \mathcal{L}(\mathcal{C}_{\mathcal{Q}}) : \forall \alpha > 0, J(\alpha X) < \infty\}. \end{aligned}$$

One may expect the functional J to be, upon identifications, the conjugate of I. We remark that the definition of J extends the one we used in the dominated case, and this holds also for the L and E spaces above, so we keep the notation from the dominated case.

Let us prove the following result, for which it is important that we work on L_I only:

Lemma 2.6.1 *If $\mathbb{Q} \in L_I\backslash\{0\}$ we have $\mathcal{P}(\mathbb{Q}) \neq \emptyset$ and the infimum defining $I(\mathbb{Q})$ is attained. Furthermore, the set L_I is linear and the functional $I(\cdot)$ is convex on it. Finally, I is l.s.c. on any segment of L_I (equipped with the embedded euclidean topology).*

Proof. Because for any $\alpha > 0$ we have that

$$\frac{\mathrm{d}\mathcal{P}(\alpha\mathbb{Q})}{\mathrm{d}|\alpha\mathbb{Q}|} = \alpha^{-1}\frac{\mathrm{d}\mathcal{P}(\mathbb{Q})}{\mathrm{d}|\mathbb{Q}|},$$

we see that $\mathbb{Q} \in L_I\backslash\{0\} \Rightarrow \mathcal{P}(\mathbb{Q}) \neq \emptyset$. The fact that $I(\mathbb{Q})$ is attained now follows from the closedness, convexity and boundedness in $L^0(|\mathbb{Q}|)$ of the set $\mathcal{Y}_{\mathbb{Q}}(1)$, upon which the corresponding infimum is computed, a Komlos-type argument and the lower semicontinuity of $Z \mapsto \mathbb{E}^{|\mathbb{Q}|}[V(Z)]$.

Let us initially prove that L_I is convex on the non-negative elements of $\mathcal{M}(\mathcal{C}_{\mathcal{Q}})$, which is clearly a convex set. So we take $\mathbb{Q}^1, \mathbb{Q}^2$ therein. If either $\mathcal{P}(\mathbb{Q}^i)$ is empty the claim is trivial. We first then suppose $\mathcal{P}(\mathbb{Q}^i)$, $i = 1, 2$, non-empty and both $\mathbb{Q}$'s not identically null, and take $\lambda \in (0, 1)$. Let us take m^i in the $L^0(\mathbb{Q}^i)$ closures of $\frac{d\mathcal{P}(\mathbb{Q}^i)}{d\mathbb{Q}^i}$ such that $I(\mathbb{Q}^i) = \mathbb{E}^{\mathbb{Q}^i}\left[V(m^i)\right]$. Thus we get:

$$\lambda I(\mathbb{Q}^1) + (1-\lambda) I(\mathbb{Q}^2) = \lambda \mathbb{E}^{\mathbb{Q}^1}\left[V(m^1)\right] + (1-\lambda)\mathbb{E}^{\mathbb{Q}^2}\left[V(m^2)\right].$$

Let us call Δ the above quantity. We observe that $\mathbb{Q} := \lambda\mathbb{Q}^1 + (1-\lambda)\mathbb{Q}^2 \gg \mathbb{Q}^1, \mathbb{Q}^2$, since we chose the latter to be measures, and so:

$$\begin{aligned}
\Delta &= \mathbb{E}^{\mathbb{Q}}\left[\lambda \frac{d\mathbb{Q}^1}{d\mathbb{Q}} V\left(\frac{m^1 d\mathbb{Q}^1/d\mathbb{Q}}{d\mathbb{Q}^1/d\mathbb{Q}}\right) + (1-\lambda)\frac{d\mathbb{Q}^2}{d\mathbb{Q}} V\left(\frac{m^2 d\mathbb{Q}^2/d\mathbb{Q}}{d\mathbb{Q}^2/d\mathbb{Q}}\right)\right] \\
&\geq \mathbb{E}^{\mathbb{Q}}\left[\left(\lambda d\mathbb{Q}^1/d\mathbb{Q} + (1-\lambda) d\mathbb{Q}^2/d\mathbb{Q}\right) V\left(\frac{\lambda m^1 d\mathbb{Q}^1/d\mathbb{Q} + (1-\lambda) m^2 d\mathbb{Q}^2/d\mathbb{Q}}{\lambda d\mathbb{Q}^1/d\mathbb{Q} + (1-\lambda) d\mathbb{Q}^2/d\mathbb{Q}}\right)\right] \\
&= \mathbb{E}^{\mathbb{Q}}\left[V\left(\lambda m^1 d\mathbb{Q}^1/d\mathbb{Q} + (1-\lambda) m^2 d\mathbb{Q}^2/d\mathbb{Q}\right)\right] \\
&=: \kappa,
\end{aligned}$$

where the first inequality comes from the joint convexity of $(x, y) \mapsto xV(y/x)$ and the equality after it is just the definition of $\mathbb{Q}$. We next easily see that $\kappa \geq I(\mathbb{Q})$, thanks to Bayes' formula, and arrive at the convexity inequality.

We now tackle the case when say $\mathbb{Q}^2 = 0$ and $\mathbb{Q}^1$ is such that $\mathcal{P}(\mathbb{Q}^1) \neq \emptyset$. For future reference we do not assume here that $\mathbb{Q}^1$ is non-negative. We have:

$$I(\lambda\mathbb{Q}^1) = \inf_{\lambda Z \in d\mathcal{P}(|\mathbb{Q}^1|)/d|\mathbb{Q}^1|} \mathbb{E}^{|\mathbb{Q}^1|}\left[\lambda V(Z)\right] = \inf_{Z \in d\mathcal{P}(|\mathbb{Q}^1|)/d|\mathbb{Q}^1|} \mathbb{E}^{|\mathbb{Q}^1|}\left[\lambda V(Z/\lambda)\right] \leq \lambda I(\mathbb{Q}^1), \tag{2.6.1}$$

where for the inequality we used that $1/\lambda \geq 1$ and $V(\cdot)$ is decreasing. Since this holds for every $\lambda \in (0, 1]$, we also get the convexity inequality since $I(\mathbb{Q}^2) = 0$. All in all, I is then convex on the non-negative elements of $\mathcal{M}(\mathcal{C}_{\mathcal{Q}})$.

The previous proves that the set of non-negative elements in L_I, which we denote L_I^+, is convex. Indeed, if $I(\alpha\mathbb{Q}^1), I(\beta\mathbb{Q}^2) < \infty$ then the convexity of I implies that $I\left(\{\lambda/\alpha + (1-\lambda)/\beta\}^{-1}[\lambda\mathbb{Q}^1 + (1-\lambda)\mathbb{Q}^2]\right) < \infty$. Thus I restricted to L_I^+ is convex. Let us now observe that if $\mathbb{Q}^1, \mathbb{P} \in L_I^+$ then $\lambda \in [0, 1] \mapsto I(\lambda\mathbb{Q}^1 + (1-\lambda)\mathbb{P})$ is l.s.c. This is true since $\lambda\mathbb{Q}^1 + (1-\lambda)\mathbb{P} \ll \frac{1}{2}(\mathbb{Q}^1 + \mathbb{P})$ and so we may rewrite our functional I on the segment between $\mathbb{Q}$ and $\mathbb{P}$ as the I that we had in the dominated case, with $\frac{1}{2}(\mathbb{Q}^1 + \mathbb{P})$ as reference measure, and so the statement reduces to the $L^0(1/2(\mathbb{Q}^1 + \mathbb{P}))$-lower semicontinuity of the restricted functional.

We will now observe that I on L_I^+ is increasing w.r.t. the order $\mathbb{Q} \leq \mathbb{P} \iff \mathbb{Q}(A) \leq \mathbb{P}(A), \forall A \in \mathcal{F}_T$. We may assume that $\mathbb{Q} \neq 0$ and so $\mathcal{P}(\mathbb{Q}) \neq \emptyset$. If we define $\mathbb{Q}^\lambda = \lambda\mathbb{Q} + (1-\lambda)\mathbb{P}$ we then have that for $\lambda \in [0, 1)$ $\mathbb{Q}^\lambda$ is equivalent to $\mathbb{P}$. Thus $\mathcal{Y}_{\mathbb{Q}^\lambda} = \mathcal{Y}_{\mathbb{P}} \times d\mathbb{P}/d\mathbb{Q}^\lambda$

and so we deduce that

$$\begin{aligned} I(\mathbb{Q}^\lambda) &= \inf_{Y \in \mathcal{Y}_{\mathbb{P}} \mathrm{d}\mathbb{P}/\mathrm{d}\mathbb{Q}^\lambda} \mathbb{E}^{\mathbb{Q}^\lambda}[V(Y)] \\ &= \inf_{Y \in \mathcal{Y}_{\mathbb{P}}} \mathbb{E}^{\mathbb{P}}\left[\frac{\mathrm{d}\mathbb{Q}^\lambda}{\mathrm{d}\mathbb{P}} V\left(\frac{Y}{\mathrm{d}\mathbb{Q}^\lambda/\mathrm{d}\mathbb{P}}\right)\right] \\ &\leq I(\mathbb{P}), \end{aligned}$$

where for the inequality we used that $\mathrm{d}\mathbb{Q}^\lambda/\mathrm{d}\mathbb{P} \leq 1$ (since $\mathbb{Q} \leq \mathbb{P}$) and that $z \mapsto zV(l/z)$ is increasing. Now taking $\lambda \nearrow 1$ and using the l.s.c. obtained previously we get that $I(\mathbb{Q}) \leq I(\mathbb{P})$.

We proceed to prove the convexity of I on the whole of L_I. Observing that I on L_I is just the composition of $\mathbb{Q} \mapsto |\mathbb{Q}|$ and the restriction of I to L_I^+, we have:

$$I(\lambda Q^1 + (1-\lambda)\mathbb{Q}^2) = I(|\lambda Q^1 + (1-\lambda)\mathbb{Q}^2|) \leq I(\lambda |Q^1| + (1-\lambda)|\mathbb{Q}^2|) \leq \lambda I(|Q^1|) + (1-\lambda) I(|\mathbb{Q}^2|),$$

and the last term equals $\lambda I(Q^1) + (1-\lambda) I(\mathbb{Q}^2)$, finally proving the convexity. The proof of the linearity of L_I follows from this with analogous arguments as in Lemma 2.5.1, and the lower semicontinuity of I on segments of lines in L_I is a trivial extension of the previous argument on segments in the positive cone. ■

Remark 2.6.1 *We stress that I need not be convex on the whole of $\mathcal{M}(\mathcal{C}_{\mathcal{Q}})$, but only in its positive cone, and this cannot be remedied by any redefinition of I on the measures for which $\mathcal{P}(\mathbb{Q}) = \emptyset$, which is what causes the trouble. The reason why convexity does appear when restricting I to L_I is that for the non-trivial measures therein $\mathcal{P}(\mathbb{Q}) \neq \emptyset$ automatically holds. This issue of course never arises in the dominated case, as one only needs to keep track of $\mathcal{P}(\mathbb{P})$ for one reference measure $\mathbb{P}$.*

We now proceed to summarize some more or less direct results related to the functionals and spaces introduced so far. We assume in the following that

$$\mathcal{Q} \subset L_I,$$

holds.

Proposition 2.6.1

- *J is a convex modular on either modular (in particular linear) spaces L_J or E_J.*
- *I is a convex modular on either modular (in particular linear) spaces L_I or E_I.*
- *$\forall \mathbb{Q} \in L_I$ the function $\alpha \in \mathbb{R}_+ \mapsto I(\alpha\mathbb{Q})$ is increasing and we have $I(\mathbb{Q}) \geq |\mathbb{Q}|(\Omega) V\left(\frac{1}{|\mathbb{Q}|(\Omega)}\right)$.*
- *A Young-type inequality holds between I and J, namely:*

$$\forall (Q, X) \in L_I \times \mathcal{L}(\mathcal{C}_{\mathcal{Q}}) \text{ holds: } I(\mathbb{Q}) + J(X) \geq \mathbb{E}^{\mathbb{Q}}(X). \tag{2.6.2}$$

Proof. The convexity of J is a consequence of it being a supremum of convex functions, and that of I has already been established in Lemma 2.6.1. The linearity of the spaces L_I, E_I, L_J, E_J is an easy consequence of the convexity of I and J (see Lemma 2.5.1).

Equation (2.6.1) implies the increasingness of $\alpha \in \mathbb{R}_+ \mapsto I(\alpha\mathbb{Q})$, and to prove the lower bound for I we simply observe that $(|\mathbb{Q}|(\Omega))^{-1}|\mathbb{Q}|$ is a probability measure and so by Jensen's inequality

$$I(\mathbb{Q}) \geq |\mathbb{Q}|(\Omega) \inf_{Z \in \mathrm{d}\mathcal{P}(|\mathbb{Q}|)/\mathrm{d}|\mathbb{Q}|} V\left(\frac{\mathbb{E}^{|\mathbb{Q}|}[Z]}{|\mathbb{Q}|(\Omega)}\right) = |\mathbb{Q}|(\Omega) V\left(\frac{1}{|\mathbb{Q}|(\Omega)}\right).$$

We now prove that L_I and L_J are modular spaces with convex modulars I and J respectively. Recall the corresponding definition 2.5.1. For I first. Axioms (1), (2) and (3) therein hold by definition, and (5) has already been established. For (4) we invoke the lower bound on I already obtained. Finally, for axiom (6), the increasingness of I along lines shows that $I(\mathbb{Q}) \geq \sup\limits_{0\leq\xi<1} I(\xi\mathbb{Q}) =: \zeta$. Now, take $\epsilon_n \nearrow 1$ so $\zeta = \lim I(\epsilon_n\mathbb{Q})$. Because I is l.s.c. along segments of lines, thanks to Lemma 2.6.1, we deduce that $\lim I(\epsilon_n\mathbb{Q}) \geq I(\mathbb{Q})$ and thus $I(\mathbb{Q}) = \zeta$ as desired. Now for J. Axioms (1), (2) and (3) are direct. If $J(\xi X) = 0$ this means $YU^{-1}(\xi X) = 0$, $|\mathbb{Q}|$-a.e. $\forall \mathbb{Q} \in L_I, Y \in \mathcal{Y}_{\mathbb{Q}}$. Notice that for $\mathbb{Q} \in L_I \backslash \{0\}$ there must exist a strictly positive $Y \in \mathcal{Y}_{\mathbb{Q}}$ and thus $U^{-1}(\xi X) = 0$, $|\mathbb{Q}|$-a.e. $\forall \mathbb{Q} \in L_I \backslash \{0\}$, implying that $U^{-1}(\xi X) = 0$ also outside a polar set, by virtue of our assumption that $\mathcal{Q} \subset L_I$. Hence $X = 0$ outside a polar set. Lastly, by increasingness of U^{-1} it holds that for fixed $Y \in \mathcal{Y}_{\mathbb{Q}}$: $YU^{-1}(\xi X) \nearrow YU^{-1}(X)$ $|\mathbb{Q}|$-a.e. as $\xi \nearrow 1$. By monotone convergence then $\mathbb{E}^{|\mathbb{Q}|}[YU^{-1}(\xi X)] \nearrow \mathbb{E}^{|\mathbb{Q}|}[YU^{-1}(X)]$ and thus $\sup_{0\leq\xi<1} \mathbb{E}^{|\mathbb{Q}|}[YU^{-1}(\xi X)] = \mathbb{E}^{|\mathbb{Q}|}[YU^{-1}(X)]$ and now taking supremum over $Y \in \mathcal{Y}_{\mathbb{Q}}$ and $\mathbb{Q} \in L_I$ we get axiom (6).

We conclude with the Young-type inequality. Fix $\mathbb{Q} \in L_I, X \in L_J$ and $Y \in \mathcal{Y}_{\mathbb{Q}}$. Then we have:

$$\begin{aligned} \mathbb{E}^{|\mathbb{Q}|}\left[V(Y)\right] + \mathbb{E}^{|\mathbb{Q}|}\left[YU^{-1}(|X|)\right] &= \mathbb{E}^{|\mathbb{Q}|}\left[V(Y) + YU^{-1}(|X|)\right] \\ &\geq \mathbb{E}^{|\mathbb{Q}|}(|X|) \\ &\geq \mathbb{E}^{\mathbb{Q}}(X), \end{aligned}$$

where the first inequality follows from the fact that $V(y) = \sup_{z>0}[z - yU^{-1}(z)]$ and the second one is clear. If we bound now $\mathbb{E}^{|\mathbb{Q}|}\left[YU^{-1}(|X|)\right]$ from above by $J(X)$ and take infimum over $Y \in \mathcal{Y}_{\mathbb{Q}}$ in $\mathbb{E}^{|\mathbb{Q}|}\left[V(Y)\right] + J(X) \geq \mathbb{E}^{\mathbb{Q}}(X)$, we arrive to the desired result. ■

Since we are dealing with convex modulars, we endow now L_I (respect. L_J) with either the Amemiya $|\cdot|_I^a$ or the equivalent Luxemburg $|\cdot|_I^\ell$ norms (respect. $|\cdot|_J^a$, $|\cdot|_J^\ell$) as in the previous sections (see e.g. (2.5.2)).

Corollary 2.6.1 *For every* $(\mathbb{Q}, X) \in L_I \times \mathcal{L}(\mathcal{C}_{\mathcal{Q}})$ *we have:*

$$I\left(\frac{\mathbb{Q}}{|\mathbb{Q}|_I^\ell}\right) \le 1 \ , \ J\left(\frac{X}{|X|_J^\ell}\right) \le 1 \text{ and further } \left|\mathbb{E}^{\mathbb{Q}}[X]\right| \le |X|_J^i |\mathbb{Q}|_I^j \le 2|X|_J^k|\mathbb{Q}|_I^k,$$

where $i, j, k \in \{a, \ell\}$ *and* $i \neq j$.

Proof. Let $\beta_n \searrow |\mathbb{Q}|_I^\ell$ so that $I(\mathbb{Q}/\beta_n) \le 1$. Then Lemma 2.6.1 shows that the inequality is preserved in the limit. The inequality for J is a trivial adaptation of that in Lemma 2.5.2. Finally the proof of the Hölder inequalities are consequence of Young's inequality in Proposition 2.6.1, along the same lines as in Proposition 2.5.3. ■

These are all encouraging results that show that probably the modular functionals and modular spaces introduced should be meaningful in the present, non-dominated, context.

We close the section with the following *open questions*:

1. Are L_I, E_I, L_J, E_J Banach spaces?
2. Is there a condition, in the spirit of Assumption 9, guaranteeing an isometric identification between $(E_I)^*$ and L_J, or at least making it possible to apply a Banach-Alaoglu-Bourbaki theorem for general dual pairings?
3. Does it hold that $J(C) \le x \iff U^{-1}(|C|) \le X_T$ for some $X \in \mathcal{X}(x)$?

Regarding these open questions, the first one seems to be the more tractable one; it is probably just an elaboration of the proof given in the dominated case. The second question seems to be far more difficult than its dominated counterpart and its truthfulness is uncertain; for instance, just talking about the norm dual of the space of finite variation measures is a very tricky subject (see e.g. Mauldin [1973], [MacNerney, 1980, Chapter II], Gordon [1966]). The non-trivial implication in the third point is very likely another hard question. If the three previous questions had an affirmative answer, this might yield existence of optimal strategies and minimax equality even in the case when $\mathcal{Q}$ is not weakly-compact.

2.7 Concluding remarks

In this work we investigated conditions on the elements of the robust utility maximization problem allowing us to relax the usual compactness assumptions on the set of densities in the uncertainty set. This was done by identifying relevant Banach spaces where a fortiori *worst-case* measures (if they exist) should live, and formulating conditions on them for the solvability of the original problem. In complete markets the relevant space is an Orlicz space. Upon granting its reflexivity, which we could do under simple growth assumptions on the utility function and its conjugate, we proved attainability of optimal strategies and existence of a worst-case measure. Furthermore, by means of entropy minimization techniques we gave an explicit characterization of this measure in terms of the solution to a certain bi-dual problem, which may be easier to solve. In particular, in many practical cases of interest, that problem is finite dimensional. For general markets, we showed that the relevant Banach space is a certain Modular space which, no matter the ingredients of the problem, is practically never (beyond the complete case) a reflexive space. Nevertheless, we could obtain in that general setting a minimax equality and the existence of optimal strategies, without resorting to model compactness assumptions nor ensuring existence of a worst-case measure. The core of the argument is that the image through the utility function of the terminal wealths live in an appropriate weak-star compact set. We argue that the Modular spaces we have introduced, or relevant modifications thereof, provide a framework where related problems can be further investigated. In particular we show how the approach simplifies some aspects of the non-robust variant of the problem as well as hints how to tackle the non-dominated, robust, version of it.[1]

[1] We would like to thank Christian Léonard for useful and stimulating exchanges about his works on entropy minimization, Freddy Delbaen for kindly commenting on the link between compactness of martingale measure densities and incompleteness, Alexander Schied for his encouraging interest on an early version of this work and for pointing out some references, Michael Kupper for providing in the beginning useful insight into the incomplete setting, Andrzej Kozek for facilitating electronic versions of his works on Orlicz-Muzielak spaces, and Ludovic Tangpi for his interest and comments on the section on the nondominated case.

3 Sensitivity analysis in optimal control

3.1 Introduction

In the present chapter we shall develop, mainly in the context of stochastic optimal control problems, a first-order sensitivity analysis for the optimal value of such problems with respect to the several parameters defining it. From optimization theory me know that the key to such analysis is to identify the Lagrange multipliers associated to the various constraints in the problem. For that matter we build a Hilbertian framework, by considering a space of Itô processes and identifying a scalar product therein, and we will be able to translate the original stochastic control problem into the language of infinite-dimensional optimization theory and then prove a correspondence between the associated Lagrange multipliers of the problem and the (weak) Pontryagin multipliers coming from Prontryiagin's principle. With such a correspondence we endeavour our sensitivity analysis, in the case of convex problems with additive perturbations and quadratic problems (Mean-Variance and LQ) with multiplicative perturbations.

The chapter is structured as follows. In Section 3.2 we introduce relevant notation and present the mentioned Hilbert space topology in the space of Itô processes, along with some needed technical lemmata. Next in Section 3.3 we identify some operators that will be of importance in the following sections and find their adjoints in terms of associated BSDEs. In Section 3.4 we define the stochastic optimal control problem, study the differentiability properties of the several functionals appearing in the data and culminate by establishing the one-to-one relationship between Lagrange and weak-Pontryagin multipliers. Then in Section 3.5 we take advantage of the Lagrange point of view and analyze the differentiability properties of the value function with respect to its parameters in the case of linear perturbations (Section 3.5.1) of convex problems, the case of stochastic Linear-Quadratic problems (Section 3.5.2) and Mean-Variance portfolio optimization problem (Section 3.5.3). The chapter is then closed with section 3.6, where we discuss in a restricted setting how a sensitivity analysis of utility maximization problems might look like, making a connection with Chapter 2 and using some ideas thereof. This chapter is a collaboration with Francisco Silva of the Université de Limoges, and except for the final section the work is based on Backhoff and Silva.

3.2 Preliminaries and functional framework

Let $T > 0$ and consider a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, on which a d-dimensional ($d \in \mathbb{N}^*$) Brownian motion $W(\cdot)$ is defined. We suppose that $\mathbb{F} = \{\mathcal{F}_t\}_{0 \leq t \leq T}$ is the natural filtration, augmented by all $\mathbb{P}$-null sets in $\mathcal{F}$, associated to $W(\cdot)$. We recall that

$\mathbb{F}$ is right-continuous. Given $\beta, p \in [1,\infty]$ and $n \in \mathbb{N}$ let us consider the Banach spaces

$$(L_{\mathbb{F}}^{\beta,p})^n \quad := \left\{ v \in L^{\beta}\left(\Omega; L^p\left([0,T];\mathbb{R}^n\right)\right) : \ (t,\omega) \to v(t,\omega) \text{ is } \mathbb{F}\text{-progressively measurable} \right\}.$$

We shall write v, $v(t)$, $v(\omega)$, $v(t,\omega)$ at will and call $\|\cdot\|_{\beta,p}$ the natural norms:

$$\|v\|_{\beta,p} := \left[\mathbb{E}\left(\|v(\omega)\|_{L^p([0,T])}^{\beta} \right) \right]^{\frac{1}{\beta}} \text{ and } \quad \|v\|_{\infty,p} := \operatorname*{ess\,sup}_{\omega\in\Omega} \|v(\omega)\|_{L^p([0,T])}.$$

The case $\beta = p = 2$ is of particular interest since $(L_{\mathbb{F}}^{2,2})^n$ is a Hilbert space endowed with the scalar product

$$\langle v_1, v_2\rangle_{L^2} := \mathbb{E}\left(\int_0^T v_1(t)^{\top} v_2(t) \mathrm{d}t \right).$$

We set $(\mathcal{M}_c^2)^n$ for the set consisting of $\mathbb{F}$-adapted, $\mathbb{R}^n$-valued square integrable martingales $x(\cdot)$ satisfying that $x(0) = 0$. Recall that in the brownian filtration $\mathbb{F}$, every martingale admits a version having $\mathbb{P}$-almost surely (a.s.) continuous trajectories (see [Revuz and Yor, 1999, Theorem 3.5, Chapter V]). In particular, the elements in $(\mathcal{M}_c^2)^n$ can be identified with $\mathbb{F}$-progressively measurable processes. Let us also recall that for every $x \in (\mathcal{M}_c^2)^n$, the martingale representation theorem (see e.g. [Ikeda and Watanabe, 1989, Chapter 2, Theorem 6.6]) provides the existence of a unique $x_2 \in (L_{\mathbb{F}}^{2,2})^{n\times d}$ such that

$$x(t) = \int_0^{\cdot} x_2(s)\mathrm{d}W(s) \qquad \forall t \in [0,T], \tag{3.2.1}$$

where, denoting $x_2^{ij} := (x_2^j)^i$,

$$\left(\int_0^{\cdot} x_2(s)\mathrm{d}W(s) \right)^i := \sum_{j=1}^{d} \int_0^{\cdot} x_2^{ij}(s)\mathrm{d}W^j(s) \qquad \text{for all } i = 1,\dots,n.$$

Note that relation (3.2.1), Doob's inequality and the Itô-isometry for the stochastic integral imply that, endowed with the scalar product

$$\langle x, y\rangle_{\mathcal{M}_c^2} := \mathbb{E}\left(x(T)^{\top} y(T) \right),$$

the set $(\mathcal{M}_c^2)^n$ is a Hilbert space which is a closed subspace of $(L_{\mathbb{F}}^{2,\infty})^n$. We now consider a larger Hilbert space, called Itô space, which is fundamental in the rest of the chapter. In order to provide a rigorous definition let us consider the application $I : \mathbb{R}^n \times (L_{\mathbb{F}}^{2,2})^n \times (L_{\mathbb{F}}^{2,2})^{n\times d} \to (L_{\mathbb{F}}^{2,\infty})^n$ defined as

$$I(x_0, x_1, x_2)(\cdot) := x_0 + \int_0^{\cdot} x_1(s)\mathrm{d}s + \int_0^{\cdot} x_2(s)\mathrm{d}W(s). \tag{3.2.2}$$

We have

Lemma 3.2.1 *The application I is well defined, injective and there exists a $c > 0$ such* *that*

$$\|I(x_0,x_1,x_2)\|_{2,\infty} \le c\left(|x_0| + \|x_1\|_{2,2} + \sum_{j=1}^{d} \|x_2^j\|_{2,2}\right). \tag{3.2.3}$$

Proof. There exists a constant $c>0$ such that for all $t \in [0,T]$,

$$|x(t)|^2 = \left|x_0 + \int_0^t x_1 \mathrm{d}s + \int_0^t x_2 \mathrm{d}W\right|^2 \le c\left(|x_0|^2 + \left|\int_0^t x_1 \mathrm{d}s\right|^2 + \left|\int_0^t x_2 \mathrm{d}W\right|^2\right).$$

By Jensen inequality applied to the first integral we get the existence of $c'>0$ such that

$$\sup_{t\in[0,T]} |x(t)|^2 \le c'\left(|x_0|^2 + \int_0^T |x_1(s)|^2 \,\mathrm{d}s + \sup_{t\in[0,T]}\left|\int_0^t x_2(s)\mathrm{d}W(s)\right|^2\right).$$

Taking the expected value, Doob's inequality and the Itô-isometry property for the stochastic integrals yields to (3.2.3). Finally, since the only continuous martingales with finite-variation are the constants (see e.g. [Revuz and Yor, 1999, Proposition 1.2]), we see that I is injective. ■

We consider the space $\mathcal{I}^n$ of $\mathbb{R}^n$-valued Itô processes defined by

$$\mathcal{I}^n := I\left(\mathbb{R}^n \times (L^{2,2}_{\mathbb{F}})^n \times (L^{2,2}_{\mathbb{F}})^{n\times d}\right).$$

By Lemma 3.2.1 have that $\mathcal{I}^n$ is a linear space which can be identified with $\mathbb{R}^n \times (L^{2,2}_{\mathbb{F}})^n \times (L^{2,2}_{\mathbb{F}})^{n\times d}$. For $x\in\mathcal{I}^n$ we set $(x_0,x_1,x_2) = I^{-1}(x)$ and we define the scalar product

$$\begin{aligned}
\langle x,y\rangle_{\mathcal{I}} &:= x_0^\top y_0 + \langle x_1,y_1\rangle_{L^2} + \textstyle\sum_{j=1}^d \langle x_2^j, y_2^j\rangle_{L^2} \\
&= x_0^\top y_0 + \langle x_1,y_1\rangle_{L^2} + \textstyle\sum_{j=1}^d \left\langle \int_0^\cdot x_2^j(s)\mathrm{d}W^j(s), \int_0^\cdot y_2^j(s)\mathrm{d}W^j(s)\right\rangle_{\mathcal{M}_c^2} \\
&= x_0^\top y_0 + \mathbb{E}\left(\int_0^T x_1(t)^\top y_1(t)\mathrm{d}t\right) + \mathbb{E}\left(\int_0^T \mathrm{tr}\left[x_2(t)^\top y_2(t)\right]\mathrm{d}t\right),
\end{aligned} \tag{3.2.4}$$

for every $x,y\in\mathcal{I}^n$, and we define the norm $\|x\|_{\mathcal{I}} := \sqrt{\langle x,x\rangle_{\mathcal{I}}}$.

Lemma 3.2.2 *The space $(\mathcal{I}^n, \|\cdot\|_{\mathcal{I}})$ is a Hilbert space which is continuously embedded in $(L^{2,\infty}_{\mathbb{F}})^n$.*

Proof. The result is a direct consequence of the fact that $\mathbb{R}^n \times (L^{2,2}_{\mathbb{F}})^n \times (L^{2,2}_{\mathbb{F}})^{n\times d}$ is a Hilbert space and Lemma 3.2.1. ■

Remark 3.2.1 (i) *We can thus identify $\mathcal{I}^n$ with $\mathbb{R}^n \times (L^{2,2}_{\mathbb{F}})^n \times (L^{2,2}_{\mathbb{F}})^{n\times d}$ and by* (3.2.1) *with $\mathbb{R}^n \times (L^{2,2}_{\mathbb{F}})^n \times (\mathcal{M}_c^2)^n$.*

(ii) *We will identify the topological dual $\mathcal{I}^*$ with the space $\mathcal{I}$ itself.*

3.3 Adjoint operators and Backward Stochastic Differential Equations (BSDEs)

We start with two basic well-known results. However, since the proofs are short, we provide the details for the reader's convenience.

Lemma 3.3.1 *Let $x \in (L^{2,\infty}_{\mathbb{F}})^n$ and $r \in (L^{2,2}_{\mathbb{F}})^n$. Then, for every $j = 1, \dots, d$,*

$$M^j(\cdot) := \int_0^{\cdot} x(s)^\top r(s) \mathrm{d}W^j(s) \quad \text{is a martingale.}$$

Proof. Since $x \in (L^{2,\infty}_{\mathbb{F}})^n$ and $r \in (L^{2,2}_{\mathbb{F}})^n$ we have that the stochastic integral M^j is well-defined and is a local-martingale. By the Burkholder-Davis-Gundy inequality (see e.g. Karatzas and Shreve [1991]) we have the existence of a constant $K > 0$ such that

$$\mathbb{E}\left(\sup_{t \in [0,T]} |M^j(t)| \right) \leq K \mathbb{E}\left[\left(\int_0^T |x(s)^\top r(s)|^2 \mathrm{d}t \right)^{\frac{1}{2}} \right] \leq K \|x\|_{2,\infty} \|r\|_{2,2},$$

where the last inequality follows from the Cauchy-Schwarz inequality. Therefore, by [Protter, 2005, Theorem 51], we have that $M^j(\cdot)$ is a martingale with null expectation. ■

Using the above result, the following one is a straightforward consequence of Itô's lemma and Lemma 3.2.2.

Lemma 3.3.2 *Let $x, y \in \mathcal{I}^n$. Then*

$$\mathbb{E}\left(x(T)^\top y(T) \right) = x_0^\top y_0 + \mathbb{E}\left(\int_0^T \left[x(t)^\top y_1(t) + y(t)^\top x_1(t) + \sum_{j=1}^d (x_2^j(t))^\top y_2^j(t) \right] \mathrm{d}t \right).$$

Given a sigma-algebra $\mathcal{G} \subseteq \mathcal{F}$ we write $L^p_{\mathcal{G}} := L^p(\Omega, \mathcal{G}, \mathbb{P})$. The following Proposition will be useful.

Proposition 3.3.1 *Let $g \in (L^2_{\mathcal{F}_T})^n$ and $a \in (L^{2,2}_{\mathbb{F}})^n$. Then, for every $z \in \mathcal{I}^n$ we have that*

$$\begin{array}{rcl} \mathbb{E}\left(g^\top z(T) \right) & = & \left\langle \mathbb{E}\left(g | \mathcal{F}_{(\cdot)} \right) + \int_0^{\cdot} \mathbb{E}(g|\mathcal{F}_t) \mathrm{d}t, z \right\rangle_{\mathcal{I}}, \\ \mathbb{E}\left(\int_0^T a(t)^\top z(t) \mathrm{d}t \right) & = & \left\langle \mathbb{E}(\int_0^T a(t) \mathrm{d}t | \mathcal{F}_{(\cdot)}) + \int_0^{\cdot} \mathbb{E}(\int_t^T a(s) \mathrm{d}s | \mathcal{F}_t) \mathrm{d}t, z \right\rangle_{\mathcal{I}}. \end{array} \tag{3.3.1}$$

In particular,

$$\mathbb{E}\left(g^\top z(T) + \int_0^T a(t)^\top z(t) \mathrm{d}t \right) = \left\langle p(0) + \int_0^{\cdot} p(t) \mathrm{d}t + \int_0^{\cdot} q(t) \mathrm{d}W(t), z \right\rangle_{\mathcal{I}}, \tag{3.3.2}$$

where $(p,q) \in \mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^{n\times d}$ *is the unique solution of the BSDE*

$$\begin{array}{rcl} \mathrm{d}p & = & -a(t)\mathrm{d}t + q(t)\mathrm{d}W(t), \\ p(T) & = & g. \end{array}$$

Proof. Let us first prove (3.3.1). Let us denote by r_g for the unique element in $(L^{2,2}_{\mathbb{F}})^{n\times d}$ (see (3.2.1)) such that

$$\mathbb{E}\left(g|\mathcal{F}_{(\cdot)}\right) = \mathbb{E}(g) + \int_0^{\cdot} r_g(t)\mathrm{d}W(t). \tag{3.3.3}$$

Lemma 3.3.2 implies that

$$\begin{array}{rcl} \mathbb{E}\left(g^\top z(T)\right) & = & \mathbb{E}\left(\mathbb{E}(g|\mathcal{F}_T)^\top z(T)\right), \\ & = & \mathbb{E}\left(\mathbb{E}(g|\mathcal{F}_0)^\top z_0 + \int_0^T \left[\mathbb{E}(g|\mathcal{F}_t)^\top z_1 + \sum_{j=1}^d (r_g^j)^\top z_2^j\right]\mathrm{d}t\right), \\ & = & \left\langle \mathbb{E}(g) + \int_0^{\cdot} \mathbb{E}(g|\mathcal{F}_t)\mathrm{d}t + \int_0^{\cdot} r_g(t)\mathrm{d}W(t), z\right\rangle_{\mathcal{I}}, \end{array}$$

which, together with (3.3.3), yields to the first identity in (3.3.1). On the other hand, setting $y(\cdot) := \int_0^{\cdot} a(t)\mathrm{d}t$,

$$\mathbb{E}\left(\int_0^T a(t)^\top z(t)\mathrm{d}t\right) = \mathbb{E}\left(\int_0^T z(t)^\top \mathrm{d}y(t)\right) = \mathbb{E}\left(y(T)^\top z(T) - \int_0^T y(t)^\top z_1(t)\mathrm{d}t\right),$$

and the second identity in (3.3.1) follows from the first one. To establish (3.3.2), let $q \in (L^{2,2}_{\mathbb{F}})^{n\times d}$ be such that

$$\mathbb{E}\left(g + \int_0^T a(t)\mathrm{d}t\big|\mathcal{F}_{(\cdot)}\right) = \mathbb{E}\left(g + \int_0^T a(t)\mathrm{d}t\right) + \int_0^{\cdot} q(t)\mathrm{d}W(t),$$

and define

$$p(t) := \mathbb{E}\left(g + \int_t^T a(s)\mathrm{d}s\big|\mathcal{F}_t\right).$$

Then

$$p(t) = \mathbb{E}\left(g + \int_0^T a(s)\mathrm{d}t\big|\mathcal{F}_t\right) - \int_0^t a(s)\mathrm{d}s = p(0) - \int_0^t a(s)\mathrm{d}s + \int_0^t q(s)\mathrm{d}W(s),$$

from which the result follows. ■

For $g \in (L^{\infty,\infty}_{\mathbb{F}})^{n\times n}$ and $h = (h^j)_{j=1}^d$ with $h^j \in (L^{\infty,\infty}_{\mathbb{F}})^{n\times n}$, let us define the operators A_g, $B_h : \mathcal{I}^n \to \mathcal{I}^n$ as

$$A_g z := \int_0^{\cdot} g(s)z(s)\mathrm{d}s, \qquad B_h z := \sum_{j=1}^d \int_0^{\cdot} h^j(s)z(s)\mathrm{d}W^j(s). \tag{3.3.4}$$

Proposition 3.3.1 has the following consequence:

Corollary 3.3.1 *The following assertions hold:*

(i) *The operator A_g is continuous and its adjoint $A_g^*: \mathcal{I}^n \to \mathcal{I}^n$ is given by*

$$A_g^* r(\cdot) = \mathbb{E}\left(\int_0^T g(t)^\top r_1(t) \mathrm{d}t \big| \mathcal{F}_{(\cdot)} \right) + \int_0^\cdot \mathbb{E}\left(\int_t^T g(s)^\top r_1(s) \mathrm{d}s \big| \mathcal{F}_t \right) \mathrm{d}t \quad \forall\, r \in \mathcal{I}^n. \tag{3.3.5}$$

Moreover,

$$A_g^* r(\cdot) = p_{g,r}(0) + \int_0^\cdot p_{g,r}(t)\mathrm{d}t + \int_0^\cdot q_{g,r}(t)\mathrm{d}W(t) \qquad \forall\, r \in \mathcal{I}^n,$$

where $(p_{g,r}, q_{g,r}) \in \mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^{n\times d}$ is the unique solution to the following BSDE

$$\begin{aligned} \mathrm{d}p(t) &= -g(t)^\top r_1(t)\mathrm{d}t + q(t)\mathrm{d}W(t), \\ p(T) &= 0. \end{aligned}$$

(ii) *The operator B_h is continuous and its adjoint $B_h^*: \mathcal{I}^n \to \mathcal{I}^n$ is given by*

$$B_h^* r(\cdot) = \sum_{j=1}^d \mathbb{E}\left(\int_0^T h^j(t)^\top r_2^j(t)\mathrm{d}t \big| \mathcal{F}_{(\cdot)} \right) + \sum_{j=1}^d \int_0^\cdot \mathbb{E}\left(\int_t^T h^j(s)^\top r_2^j(s)\mathrm{d}s \big| \mathcal{F}_t \right) \mathrm{d}t \quad \forall\, r \in \mathcal{I}^n. \tag{3.3.6}$$

Moreover,

$$B_h^* r(\cdot) = p_{h,r}(0) + \int_0^\cdot p_{h,r}(t)\mathrm{d}t + \int_0^\cdot q_{h,r}(t)\mathrm{d}W(t) \qquad \forall\, r \in \mathcal{I}^n,$$

where $(p_{h,r}, q_{h.r}) \in \mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^{n\times d}$ is the unique solution to the following BSDE

$$\begin{aligned} \mathrm{d}p(t) &= -\textstyle\sum_{j=1}^d h^j(t)^\top r_2^j(t)\mathrm{d}t + q(t)\mathrm{d}W(t), \\ p(T) &= 0. \end{aligned}$$

Consequently, the adjoint of $A_g + B_h$ is given by

$$(A_g + B_h)^* r(\cdot) = p_r(0) + \int_0^\cdot p_r(t)\mathrm{d}t + \int_0^\cdot q_r(t)\mathrm{d}W(t) \qquad \forall\, r \in \mathcal{I}^n,$$

where $(p_r, q_r) \in \mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^{n\times d}$ is the unique solution to the following BSDE

$$\begin{aligned} \mathrm{d}p(t) &= -\left[g(t)^\top r_1(t) + \textstyle\sum_{j=1}^d h^j(t)^\top r_2^j(t) \right] \mathrm{d}t + q(t)\mathrm{d}W(t), \\ p(T) &= 0. \end{aligned}$$

Proof. For all $z \in \mathcal{I}^n$, we have that

$$
\begin{array}{rcl}
\|A_g z\|_{\mathcal{I}} & = & \left[\mathbb{E}\left(\int_0^T |g(t)z(t)|^2 \mathrm{d}t\right)\right]^{\frac{1}{2}} \leq n\|g\|_{\infty,\infty}\|z\|_{2,2} \leq n\sqrt{T}\|g\|_{\infty,\infty}\|z\|_{2,\infty}, \\
\|B_h z\|_{\mathcal{I}} & = & \sum_{j=1}^{d}\left[\mathbb{E}\left(\int_0^T |h^j(t)z(t)|^2 \mathrm{d}t\right)\right]^{\frac{1}{2}} \leq nd\|g\|_{\infty,\infty}\|z\|_{2,2} \leq nd\sqrt{T}\|g\|_{\infty,\infty}\|z\|_{2,\infty}.
\end{array}
$$

Therefore, Lemma 3.2.2 implies that the linear operators are indeed continuous. We also have that, by Lemma 3.3.1:

$$
\begin{array}{rcl}
\langle r, A_g z\rangle_I & = & \mathbb{E}\left(\int_0^T r_1(t)^\top g(t)z(t)\mathrm{d}t\right), \\
 & = & \left\langle \mathbb{E}(\int_0^T g(t)^\top r_1(t)\mathrm{d}t|\mathcal{F}_{(\cdot)}) + \int_0^{\cdot} \mathbb{E}(\int_t^T g(s)^\top r_1(s)\mathrm{d}s|\mathcal{F}_t)\mathrm{d}t, z\right\rangle_{\mathcal{I}},
\end{array}
$$

which, by Lemma 3.3.1, implies the expression for A_g^* in (i). The corresponding identity for B_g^* in (ii) is obtained by an analogous argument, while assertion (iii) is a direct consequence of (i)-(ii). ■

3.4 Optimal control problems and Lagrange multipliers

Let us introduce some notation and assumptions. For a differentiable function $(a, b) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \mapsto \psi(a, b) \in \mathbb{R}^{n_3}$ we denote by $\psi_a(a, b) \in \mathbb{R}^{n_3 \times n_1}$ and $\psi_b(a, b) \in \mathbb{R}^{n_3 \times n_2}$ the corresponding Jacobian matrices. Let $f : \Omega \times [0, T] \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ and $\sigma : \Omega \times [0, T] \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^{n \times d}$. In what follows we use the notation $f = (f^i)_{(1 \leq i \leq n)}$ and $\sigma = (\sigma^{ij})_{1 \leq i \leq n,\, 1 \leq j \leq d}$, where each f^i and σ^{ij} is real valued. The columns of σ are written σ^j for $j = 1, \ldots, d$. We suppose that:

Assumption 11 *The maps $\psi = f^j, \sigma^{ij}$ satisfy:*

(i) *ψ is $\mathcal{F}_T \otimes \mathcal{B}([0, T] \times \mathbb{R}^n \times \mathbb{R}^m)$-measurable.*
(ii) *For a.a. $(\omega, t) \in \Omega \times [0, T]$ the mapping $(x, u) \mapsto \psi(\omega, t, x, u)$ is C^1, the application $(\omega, t) \in \Omega \times [0, T] \mapsto \psi(\omega, t, \cdot, \cdot) \in C^1(\mathbb{R}^n \times \mathbb{R}^m)$ is progressively measurable and there exists $c_1 > 0$ such that almost surely in (ω, t)*

$$
\left\{
\begin{array}{c}
|\psi(\omega, t, x, u)| \leq c_1 (1 + |x| + |u|), \\
|\psi_x(\omega, t, x, u)| + |\psi_u(\omega, t, x, u)| \leq c_1,
\end{array}
\right. \tag{3.4.1}
$$

Remark 3.4.1 *Note that under Assumption 11 for every $(x, u) \in \mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^m$ we have that $(\omega, t) \mapsto \psi(\omega, t, x(\omega, t), u(\omega, t))$ is progressively measurable, and therefore $\int_0^{\cdot} f(\omega, t, x(\omega, t), u(\omega, t))\mathrm{d}t$ and $\int_0^{\cdot} \sigma(\omega, t, x(\omega, t), u(\omega, t))\mathrm{d}W(t)$ are two a.s. continuous progressively measurable processes. The latter is also a square integrable continuous martingale.*

Let us consider the application $G : \mathcal{I}^n \times (L_{\mathbb{F}}^{2,2})^m \to \mathcal{I}^n$ defined by

$$G(x,u) := \int_0^{\cdot} f(s,x(s),u(s))\mathrm{d}s + \int_0^{\cdot} \sigma(s,x(s),u(s))\mathrm{d}W(s) - x(\cdot). \tag{3.4.2}$$

Lemma 3.4.1 *Under Assumption 11 the mapping G is Lipschitz continuous and Gâteaux differentiable. Its Gâteaux derivative $DG(x,u) : \mathcal{I}^n \times (L_{\mathbb{F}}^{2,2})^m \to \mathcal{I}^n$ is given by*

$$\begin{aligned} DG(x,u)(z,v)(\cdot) \;&=\; \textstyle\int_0^{\cdot} \left[f_x(t,x(t),u(t))z(t) + f_u(t,x(t),u(t))v(t)\right]\mathrm{d}t \\ &\quad + \textstyle\int_0^{\cdot} \left[\sigma_x(t,x(t),u(t))z(t) + \sigma_u(t,x(t),u(t))v(t)\right]\mathrm{d}W(t) - z(\cdot), \end{aligned} \tag{3.4.3}$$

for all $(z,v) \in \mathcal{I}^n \times (L_{\mathbb{F}}^{2,2})^m$. Moreover, for every u, $v \in (L_{\mathbb{F}}^{2,2})^m$ and every $x \in \mathcal{I}^n$ we have that $DG(x,u)(\cdot,v) : \mathcal{I}^n \to \mathcal{I}^n$ is bijective.

Proof. Given $z \in \mathcal{I}^n$, $v \in (L_{\mathbb{F}}^{2,2})^m$ and $\tau > 0$, by a first order Taylor expansion of f and σ we obtain

$$\begin{aligned} G(x+\tau z, u+\tau v) - G(x,u) \;&=\; \tau \textstyle\int_0^{\cdot} \left[f_x(t,x(t),u(t))z(t) + f_u(t,x(t),u(t))v(t) + r_1(t,\tau)\right]\mathrm{d}t \\ &\quad + \tau \textstyle\int_0^{\cdot} \left[\sigma_x(t,x(t),u(t))z(t) + \sigma_u(t,x(t),u(t))v(t) + r_2(t,\tau)\right]\mathrm{d}W(t) \\ &\quad - \tau z(\cdot), \end{aligned} \tag{3.4.4}$$

where

$$\begin{aligned} r_1(\omega,t,\tau) \;&:=\; \textstyle\int_0^1 \left[Df(t,x(t)+\theta\tau z(t),u(t)+\theta\tau v(t)) - Df(t,x(t),u(t))\right](z,v)\mathrm{d}\theta, \\ r_2(\omega,t,\tau) \;&:=\; \textstyle\int_0^1 \left[D\sigma(t,x(t)+\theta\tau z(t),u(t)+\theta\tau v(t)) - D\sigma(t,x(t),u(t))\right](z,v)\mathrm{d}\theta. \end{aligned}$$

By Assumption 11(ii), we have that

$$|r_1(\omega,t,\tau)|^2 + |r_2(\omega,t,\tau)|^2 \le c'\left(|z(\omega,t)|^2 + |v(\omega,t)|^2\right) \quad \text{for a.a. } (\omega,t) \in \Omega \times [0,T]. \tag{3.4.5}$$

Since the left hand side of (3.4.5) converges a.s. to 0 as $\tau \downarrow 0$, we deduce with Lemma 3.2.2 and the dominated convergence theorem that

$$\mathbb{E}\left(\int_0^T |r_1(t,\tau)|^2\mathrm{d}t\right) + \mathbb{E}\left(\int_0^T |r_2(t,\tau)|^2\mathrm{d}t\right) \to 0 \quad \text{as } \tau \downarrow 0,$$

and thus (3.4.3) follows by dividing by τ in (3.4.4), taking the limit $\tau \downarrow 0$ and the definition of convergence in $\mathcal{I}^n$. Now, fix $v \in (L_{\mathbb{F}}^{2,2})^m$ and $\xi \in \mathcal{I}^n$. Let us prove that there exists $z \in \mathcal{I}^n$ such that $DG(x,u)(z,v) = \xi$. By definition, this is equivalent to solving the SDE

$$\begin{aligned} \mathrm{d}z \;&=\; \left[f_x(t,x(t),u(t))z(t) + f_u(t,x(t),u(t))v(t) - \xi_1\right]\mathrm{d}t \\ &\quad + \left[\sigma_x(t,x(t),u(t))z(t) + \sigma_u(t,x(t),u(t))v(t) - \xi_2\right]\mathrm{d}W(t) \\ z(0) \;&=\; -\xi_0. \end{aligned}$$

Since $(\xi_1, \xi_2) \in (L^{2,2}_{\mathbb{F}})^n \times (L^{2,2}_{\mathbb{F}})^{n\times d}$, under Assumption 11 classical results for solvability of linear SDEs (see e.g. [Bismut, 1976b, Theorem 2.1]) imply that the above equation has a unique solution. ■

Remark 3.4.2 *Note that under our assumptions G is Lipschitz. Therefore, by classical results (see e.g.* [Bonnans and Shapiro, 2000, Proposition 2.49]) *we have that G is Hadamard differentiable, i.e.*

$$\lim_{\tau\to 0,(z',v')\to(z,v)} \frac{G(x+\tau z', u+\tau v')(\cdot) - G(x,u)(\cdot)}{\tau} = DG(x,u)(z,v)(\cdot) \qquad \text{in } \mathcal{I}^n.$$

In general, it is not clear whether G is C^1. However, if f and σ are affine functions of the pair (x,u), it can be easily checked that $(x,u) \in \mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^m \to DG(x,u) \in L(\mathcal{I}^n, \mathcal{I}^n)$ is continuous ($L(\mathcal{I}^n, \mathcal{I}^n)$ is the space of bounded linear applications from $\mathcal{I}^n$ to $\mathcal{I}^n$), which implies that G is continuously differentiable.

Now, let

$$\ell : \Omega\times[0,T]\times\mathbb{R}^n\times\mathbb{R}^m \to \mathbb{R}, \quad \Phi : \Omega\times\mathbb{R}^n \to \mathbb{R} \quad \Phi_E : \Omega\times\mathbb{R}^n \to \mathbb{R}^{n_E}, \quad \Phi_I : \Omega\times\mathbb{R}^n \to \mathbb{R}^{n_I}.$$

Assumption 12 *We suppose that:*
(i) *The maps ℓ and $\psi = \Phi$, Φ^i_E, Φ^j_I ($1 \leq i \leq n_E$ and $1 \leq j \leq n_I$) are respectively $\mathcal{F}_T \otimes \mathcal{B}([0,T]\times\mathbb{R}^n\times\mathbb{R}^m)$ and $\mathcal{F}_T \otimes \mathcal{B}(\mathbb{R}^n)$ measurable.*
(ii) *For a.a. (ω,t) the maps $(x,u) \mapsto \ell(\omega,t,x,u)$ and $x \mapsto \psi(\omega,x)$ are C^1. The application $(\omega,t) \in \Omega\times[0,T] \mapsto \ell(\omega,t,\cdot,\cdot) \in C^1(\mathbb{R}^n\times\mathbb{R}^m)$ is progressively measurable. In addition, there exists $c_2 > 0$ such that almost surely in (ω,t) we have that*

$$\left\{ \begin{array}{c} |\ell(\omega,t,x,u)| \leq c_2\,(1+|x|+|u|)^2\,, \\ |\ell_x(\omega,t,x,u)| + |\ell_u(\omega,t,x,u)| \leq c_2\,(1+|x|+|u|)\,, \\ |\psi(\omega,x)| \leq c_2\,(1+|x|)^2\,,\ |\psi_x(\omega,x)| \leq c_2\,(1+|x|)\,. \end{array} \right. \tag{3.4.6}$$

We define $F : \mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^m \to \mathbb{R}$, $G_E : \mathcal{I}^n \to \mathbb{R}^{n_E}$ and $G_I : \mathcal{I}^n \to \mathbb{R}^{n_I}$ as

$$\begin{array}{rcl} F(x,u) & := & \mathbb{E}\left(\int_0^T \ell(t,x(t),u(t))\mathrm{d}t + \Phi(x(T))\right), \\ G^i_E(x) & := & \mathbb{E}\left(\Phi^i_E(x(T))\right) \quad \forall\, i = 1,\ldots,n_E, \\ G^j_I(x) & := & \mathbb{E}\left(\Phi^j_I(x(T))\right) \quad \forall\, j = 1,\ldots,n_I. \end{array} \tag{3.4.7}$$

Lemma 3.4.2 *The functions F, G_E and G_I are continuously differentiable (in the Fréchet sense) and for every (x,u), $(z,v) \in \mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^m$ we have that*

$$DF(x,u)(z,v) = \mathbb{E}\left(\int_0^T [\ell_x(t,x,u)z + \ell_u(t,x,u)v]\,\mathrm{d}t + \Phi_x(x(T))z(T)\right),$$

$$\begin{aligned} DG^i_E(x,u)(z,v) &= \mathbb{E}\left((\Phi^i_E)_x(x(T))z(T)\right) \qquad \forall\, 1 \le i \le n_E,\\ DG^j_I(x,u)(z,v) &= \mathbb{E}\Big((\Phi^j_I)_x(x(T))z(T)\Big) \qquad \forall\, 1 \le j \le n_I. \end{aligned} \tag{3.4.8}$$

Proof. The proof that F is Gâteaux differentiable and that its Gâteaux derivative satisfies the first equation in (3.4.8) follows the same lines as the proof of Lemma 3.4.1. Now note that given $(z,v) \in \mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^m$ with $\|z\|_{\mathcal{I}} = \|v\|_{2,2} = 1$, for all (x,u), $(x',u') \in \mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^m$,

$$\begin{aligned} |DF(x,u)(z,v) - DF(x',u')(z,v)| \;\le\;& \|z\|_{2,\infty}\left(\mathbb{E}\left[\int_0^T |\ell_x(t,x,u)-\ell_x(t,x',u')|\,\mathrm{d}t\right]^2\right)^{\frac12}\\ &+\|z\|_{2,\infty}\left(\mathbb{E}\left[|\Phi_x(x(T))-\Phi_x(x'(T))|^2\right]\right)^{\frac12}\\ &+\|v\|_{2,2}\left(\mathbb{E}\left[\int_0^T |\ell_u(t,x,u)-\ell_u(t,x',u')|^2\,\mathrm{d}t\right]\right)^{\frac12}. \end{aligned}$$

Therefore, by Lemma 3.2.1 we get that

$$\sup_{(z,v)\in\mathcal{I}^n\times(L^{2,2}_{\mathbb{F}})^m\,,\ \|z\|_{\mathcal{I}}=\|v\|_{2,2}=1} |DF(x,u)(z,v) - DF(x',u')(z,v)|^2 \le c w(x',u'),$$

where

$$\begin{aligned} w(x',u') :=\;& \mathbb{E}\left(\int_0^T \left[|\ell_x(t,x,u)-\ell_x(t,x',u')|^2 + |\ell_u(t,x,u)-\ell_u(t,x',u')|^2\right]\mathrm{d}t\right.\\ &\left. + |\Phi_x(x(T))-\Phi_x(x'(T))|^2\right). \end{aligned}$$

Since ℓ_x, ℓ_u and Φ_x satisfy the linear growth property in (3.4.6), we have by dominated convergence that $w(x',u') \to 0$ as $\|x'-x\|_{\mathcal{I}} + \|u'-u\|_{2,2} \to 0$. Thus DF is continuous and therefore F is Fréchet differentiable. The proof of the analogous result for G_E and G_I goes similarly. ■

Let $U \subseteq \mathbb{R}^m$ be a non-empty, closed and convex set and define

$$\mathcal{U} := \left\{u \in (L^{2,2}_{\mathbb{F}})^m \;:\; u(\omega,t) \in U \text{ for a.a. } (\omega,t) \in \Omega\times[0,T]\right\}. \tag{3.4.9}$$

We consider the optimal control problem

$$\min_{x\in\mathcal{I}^n,u\in(L^{2,2}_{\mathbb{F}})^m} F(x,u) \text{ s.t. } G(x,u)+x_0=0,\;\; G_E(x)=0 \text{ and } G_I(x)\le 0,\;\; u\in\mathcal{U}. \tag{$\mathcal{SP}$}$$

Remark 3.4.3 *Usually the optimal control problem above is stated only in terms of u. Indeed, under our assumptions, for every $u \in (L^{2,2}_{\mathbb{F}})^m$ there exists a unique $x[u] \in \mathcal{I}^n$ such that $G(x[u],u)+x_0=0$. Therefore, problem $(\mathcal{SP})$ can be equivalently written as*

$$\text{Min}_{\,u}\; F(x[u],u) \text{ s.t. } G_E(x[u])=0 \text{ and } G_I(x[u])\le 0,\;\; u\in\mathcal{U}. \tag{$\mathcal{SP}'$}$$

We have preferred to consider the minimization problem in terms of the pair (x,u) and thus to maintain explicitly the constraint $G(x,u)+x_0=0$ in order to associate a

Lagrange multiplier to it.

Definition 3.4.1 (i) *We say that $(x,u) \in \mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^m$ is feasible for $(\mathcal{SP})$ if $G(x,u)+x_0=0$, $G_E(x)=0$, $G_I(x) \leq 0$ and $u \in \mathcal{U}$. The set of feasible pairs for problem $(\mathcal{SP})$ is denoted by $F(\mathcal{SP})$.*

(ii) *We say that $(\bar{x},\bar{u}) \in F(\mathcal{SP})$ is a local solution of $(\mathcal{SP})$ iff $\exists\ \varepsilon > 0$ such that $F(\bar{x},\bar{u}) \leq F(x,u)$ for all $(x,u) \in F(\mathcal{SP})$ satisfying that $\|x-\bar{x}\|_{\mathcal{I}} + \|u-\bar{u}\|_{2,2} \leq \varepsilon$.*

3.4.1 Weak-Pontryagin multipliers and Lagrange multipliers

Given $\alpha \geq 0$ the *Hamiltonian* $H[\alpha] : \Omega \times [0,T] \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^{n\times d} \to \mathbb{R}$ is defined as

$$H[\alpha](\omega,t,x,u,p,q) := \alpha\ell(\omega,t,x,u) + p^\top f(\omega,t,x,u) + \sum_{j=1}^{d} (q^j)^\top \sigma^j(\omega,t,x,u). \quad (3.4.10)$$

Definition 3.4.2 (weak-Pontryagin multiplier) *We say that $0 \neq (\bar{\alpha},\bar{p},\bar{q},\bar{\lambda}_E,\bar{\lambda}_I) \in \mathbb{R} \times \mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^{n\times d} \times \mathbb{R}^{n_E} \times \mathbb{R}^{n_I}$ is a generalized weak-Pontryagin multiplier at $(\bar{x},\bar{u})$ if*

$$\begin{array}{rcl}
\mathrm{d}\bar{p}(t) & = & -H_x[\bar{\alpha}](t,\bar{x}(t),\bar{u}(t),\bar{p}(t),\bar{q}(t))^\top \mathrm{d}t + \bar{q}(t)\mathrm{d}W(t), \\
\bar{p}(T) & = & \bar{\alpha}\Phi_x(\bar{x}(T))^\top + (\Phi_E)_x(\bar{x}(T))^\top \bar{\lambda}_E + (\Phi_I)_x(\bar{x}(T))^\top \bar{\lambda}_I, \\
0 & \leq & H_u[\bar{\alpha}](\omega,t,\bar{x}(t),\bar{u}(t),\bar{p}(t),\bar{q}(t))(v-\bar{u}(\omega,t)) \quad \forall v \in U \text{ a.a. } (\omega,t) \in \Omega \times [0,T], \\
0 & < & |\bar{\alpha}| + |\bar{\lambda}_I| + |\bar{\lambda}_E|, \\
0 & = & \bar{\lambda}_I^j G_I^j(\bar{x}(T)) \quad \forall\, j = 1,\ldots,n_I, \\
0 & \leq & \bar{\lambda}_I^j \quad \forall\, j = 1,\ldots,n_I \ \text{ and } \ 0 \leq \bar{\alpha}.
\end{array} \quad (3.4.11)$$

If $\bar{\alpha} > 0$ (and therefore can be normalized to $\bar{\alpha} = 1$), we say that $0 \neq (\bar{p},\bar{q},\bar{\lambda}_E,\bar{\lambda}_I)$ is a weak-Pontryagin multiplier at $(\bar{x},\bar{u})$. The set of weak-Pontryagin multipliers is denoted by $\Lambda_{wP}(\bar{x},\bar{u})$.

It is well known that the following stochastic *weak-Pontryagin minimum principle* holds (see e.g. Peng [1990], Mou and Yong [2007] and [Yong and Zhou, 1999, Theorem 3.2, Chapter 3])

Theorem 3.4.1 (weak-Pontryagin minimum principle) *Assume that Assumptions 11-12 hold and let $(\bar{x},\bar{u}) \in \mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^m$ be a local solution of (SP). Then, there exists at least one weak-Pontryagin multiplier at $(\bar{x},\bar{u})$.*

Remark 3.4.4 (i) *In view of Remark 3.4.3, Pontryagin principles are usually stated for a local solution $\bar{u}$ of $(\mathcal{SP}')$. However, we easily check that $\bar{u}$ is a local solution of $(\mathcal{SP}')$ if and only if $(\bar{x},\bar{u})$ is a local solution of $(\mathcal{SP})$.*

(ii) *We called the result of Theorem 3.4.1 a weak-Pontryagin minimum principle, since* *in general more information can be obtained. In fact, even when U is not convex, under a Lipschitz type assumption on the second derivatives of the data, a second pair of adjoint processes can be introduced in such a manner that the optimal $\bar{u}$ minimizes an associated Hamiltonian in U. In the particular case when U is convex,* (3.4.11) *is an easy consequence of this result* (*see e.g.* Peng [1990] *and* [Yong and Zhou, 1999, Chapter 3]).

The *Lagrangian* $\mathcal{L} : \mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^m \times \mathcal{I}^n \times \mathbb{R} \times \mathbb{R}^{n_E} \times \mathbb{R}^{n_I} \to \mathbb{R}$ associated to problem $(\mathcal{SP})$ is defined by

$$\mathcal{L}(x, u, \alpha, \lambda_{\mathcal{I}}, \lambda_E, \lambda_I) := \alpha F(x, u) + \langle \lambda_{\mathcal{I}}, G(x, u) + x_0 \rangle_{\mathcal{I}} + \lambda_E^\top G_E(x) + \lambda_I^\top G_I(x), \quad (3.4.12)$$

where $G : \mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^m \to \mathcal{I}^n$ is defined in (3.4.2) and F, G_E and G_I are defined in (3.4.7).

Definition 3.4.3 *We say that $0 \neq (\bar{\alpha}, \bar{\lambda}_{\mathcal{I}}, \bar{\lambda}_E, \bar{\lambda}_I)$ is a generalized Lagrange multiplier at $(\bar{x}, \bar{u})$ if*

$$\begin{array}{rcl} 0 & = & D_x \mathcal{L}(\bar{x}, \bar{u}, \bar{\alpha}, \bar{\lambda}_{\mathcal{I}}, \bar{\lambda}_E, \bar{\lambda}_I), \\ 0 & \leq & D_u \mathcal{L}(\bar{x}, \bar{u}, \bar{\alpha}, \bar{\lambda}_{\mathcal{I}}, \bar{\lambda}_E, \bar{\lambda}_I)(v - \bar{u}) \quad \forall v \in \mathcal{U}, \\ 0 & < & |\bar{\alpha}| + |\bar{\lambda}_I| + |\bar{\lambda}_E|, \\ 0 & = & \bar{\lambda}_I^j G_I^j(\bar{x}(T)) \quad \forall\, j = 1, \ldots, n_I, \\ 0 & \leq & \bar{\lambda}_I^j \quad \forall\, j = 1, \ldots, n_I \;\; \textit{and} \;\; 0 \leq \bar{\alpha}. \end{array} \quad (3.4.13)$$

If $\bar{\alpha} > 0$ (and therefore can be normalized to $\bar{\alpha} = 1$) we will say that $0 \neq (\bar{\lambda}_{\mathcal{I}}, \bar{\lambda}_E, \bar{\lambda}_I)$ is a Lagrange multiplier at $(\bar{x}, \bar{u})$ and we will eliminate the $\bar{\alpha}$ from the arguments of $\mathcal{L}$. The set of Lagrange multipliers is denoted by $\Lambda_L(\bar{x}, \bar{u})$.

Remark 3.4.5 *If no final constraints are present, we will eliminate (λ_E, λ_I) from the arguments of $\mathcal{L}$*

Using the theoretical framework introduced in the previous sections we can prove the following

Theorem 3.4.2 *Let $(\bar{x}, \bar{u}) \in F(\mathcal{SP})$. If $(\bar{\alpha}, \bar{p}, \bar{q}, \bar{\lambda}_E, \bar{\lambda}_I)$ is a generalized weak-Pontryagin multiplier at $(\bar{x}, \bar{u})$ then $(\bar{\alpha}, \bar{\lambda}_{\mathcal{I}}, \bar{\lambda}_E, \bar{\lambda}_I)$ is a generalized Lagrange multiplier at $(\bar{x}, \bar{u})$, where*

$$\bar{\lambda}_{\mathcal{I}}(\cdot) := \bar{p}(0) + \int_0^{\cdot} \bar{p}(s) \mathrm{d}s + \int_0^{\cdot} \bar{q}(s) \mathrm{d}W(s). \quad (3.4.14)$$

Conversely, if $(\bar{\alpha}, \bar{\lambda}_{\mathcal{I}}, \bar{\lambda}_E, \bar{\lambda}_I)$ is a generalized Lagrange multiplier at $(\bar{x}, \bar{u})$ then $(\bar{\lambda}_{\mathcal{I}})_0 = (\bar{\lambda}_{\mathcal{I}})_1(0)$ and $(\bar{\alpha}, (\bar{\lambda}_{\mathcal{I}})_1, (\bar{\lambda}_{\mathcal{I}})_2, \bar{\lambda}_E, \bar{\lambda}_I)$ is a generalized weak-Pontryagin multiplier at $(\bar{x}, \bar{u})$.

Remark 3.4.6 *If $\bar{\alpha} = 1$ we can replace in the statement of the theorem "generalized weak-Pontryagin multiplier" by "weak-Pontryagin multiplier" and "generalized Lagrange multiplier" by "Lagrange multiplier."*

Proof. For notational convenience we set $\ell_x(t) := \ell_x(t, \bar{x}(t), \bar{u}(t))$ with analogous definitions for $f_u(t)$, $\sigma_x(t)$ and $\sigma_u(t)$. Let $(\bar{\alpha}, \bar{p}, \bar{q}, \bar{\lambda}_E, \bar{\lambda}_I)$ be a generalized weak-Pontryagin multiplier at $(\bar{x}, \bar{u})$. In order to prove that $(\bar{\alpha}, \bar{\lambda}_{\mathcal{I}}, \bar{\lambda}_E, \bar{\lambda}_I)$, with $\bar{\lambda}_{\mathcal{I}}$ given by (3.4.14), is a generalized Lagrange multiplier at $(\bar{x}, \bar{u})$ it suffices to show that the first two relations in (3.4.13) hold true. For the first one, for every $z \in \mathcal{I}^n$, Lemma 3.4.1 and Lemma 3.4.2 imply that

$$\begin{array}{rcl} \bar{\alpha} D_x F(x,u)z & = & \mathbb{E}\left(\int_0^T \bar{\alpha}\ell_x(t)z(t)\mathrm{d}t + \bar{\alpha}\Phi_x(\bar{x}(T))z(T)\right), \\ \langle \bar{\lambda}_E, D_x G_E(\bar{x})z\rangle & = & \mathbb{E}\left(\bar{\lambda}_E^\top(\Phi_E)_x(\bar{x}(T))z(T)\right), \\ \langle \bar{\lambda}_I, D_x G_I(\bar{x})z\rangle & = & \mathbb{E}\left(\bar{\lambda}_I^\top(\Phi_I)_x(\bar{x}(T))z(T)\right), \\ \langle \bar{\lambda}_{\mathcal{I}}, D_x G(\bar{x})z\rangle_{\mathcal{I}} & = & \mathbb{E}\left(\int_0^T \left[\bar{\lambda}_1(t)^\top f_x(t) + \sum_{j=1}^d \bar{\lambda}_2^j(t)^\top \sigma_x^j(t)\right] z(t)\mathrm{d}t\right) - \langle \bar{\lambda}_{\mathcal{I}}, z\rangle_{\mathcal{I}}. \end{array} \tag{3.4.15}$$

Using Proposition 3.3.1, with $a = \bar{\alpha}\ell_x(t)$ and $g^\top = \bar{\alpha}\Phi_x(\bar{x}(T)) + \bar{\lambda}_E^\top(\Phi_E)_x(\bar{x}(T)) + \bar{\lambda}_I^\top(\Phi_I)_x(\bar{x}(T))$, we get, recalling (3.3.4), that

$$\begin{array}{rcl} D_x\mathcal{L}(x,u,\alpha,\lambda_{\mathcal{I}},\lambda_E,\lambda_I)z & = & \left\langle \hat{p}(0) + \int_0^\cdot \hat{p}\mathrm{d}t + \int_0^\cdot \hat{q}\mathrm{d}W - \bar{\lambda}_{\mathcal{I}}, z\right\rangle_{\mathcal{I}} + \langle \lambda_{\mathcal{I}}, (A_{f_x} + B_{\sigma_x})z\rangle_{\mathcal{I}}, \\ & = & \left\langle \hat{p}(0) + \int_0^\cdot \hat{p}\mathrm{d}t + \int_0^\cdot \hat{q}\mathrm{d}W + (A_{f_x} + B_{\sigma_x})^*\bar{\lambda}_{\mathcal{I}} - \bar{\lambda}_{\mathcal{I}}, z\right\rangle_{\mathcal{I}}, \end{array}$$

where $(\hat{p}, \hat{q}) \in \mathcal{I}^n \times (L_{\mathbb{F}}^{2,2})^{n\times d}$ is the unique solution of

$$\begin{array}{rcl} \mathrm{d}\hat{p}(t) & = & -\bar{\alpha}\ell_x(t)^\top\mathrm{d}t + \hat{q}(t)\mathrm{d}W(t), \\ \hat{p}(T) & = & \bar{\alpha}\Phi_x(\bar{x}(T))^\top + (\Phi_E)_x(\bar{x}(T))^\top\bar{\lambda}_E + (\Phi_I)_x(\bar{x}(T))^\top\bar{\lambda}_I. \end{array}$$

By Corollary 3.3.1 we get that

$$D_x\mathcal{L}(\bar{x},\bar{u},\alpha,\lambda_{\mathcal{I}},\lambda_E,\lambda_I)z = \left\langle p(0) + \int_0^\cdot p(t)\mathrm{d}t + \int_0^\cdot q(t)\mathrm{d}W(t) - \bar{\lambda}_{\mathcal{I}}, z\right\rangle_{\mathcal{I}}, \tag{3.4.16}$$

where $(p, q) \in \mathcal{I}^n \times (L_{\mathbb{F}}^{2,2})^{n\times d}$ is the unique solution of

$$\begin{array}{rcl} \mathrm{d}p(t) & = & -\left[\bar{\alpha}\ell_x(t)^\top + f_x(t)^\top(\bar{\lambda}_{\mathcal{I}})_1(t) + \sigma_x(t)^\top(\bar{\lambda}_{\mathcal{I}})_2(t)\right]\mathrm{d}t + q(t)\mathrm{d}W(t), \\ p(T) & = & \bar{\alpha}\Phi_x(\bar{x}(T))^\top + (\Phi_E)_x(\bar{x}(T))^\top\bar{\lambda}_E + (\Phi_I)_x(\bar{x}(T))^\top\bar{\lambda}_I. \end{array} \tag{3.4.17}$$

Since $((\bar{\lambda}_{\mathcal{I}})_1, (\bar{\lambda}_{\mathcal{I}})_2) = (\bar{p}, \bar{q})$, by (3.4.11) we get that $p(T) - \bar{p}(T) = 0$ and $\mathrm{d}[p - \bar{p}](t) = [q(t) - \bar{q}(t)]\mathrm{d}W(t)$ which yields to $p = \bar{p}$, $q = \bar{q}$ and in particular $p(0) = \bar{p}(0)$, hence the first relation in (3.4.13) follows from (3.4.16). In order to prove the second relation in (3.4.13) it suffices to note that for all $v \in \mathcal{U}$

$$D_u\mathcal{L}(\bar{x},\bar{u},\bar{\alpha},\bar{\lambda}_{\mathcal{I}},\bar{\lambda}_E,\bar{\lambda}_I)(v-\bar{u}) = \mathbb{E}\left(\int_0^T H_u[\bar{\alpha}](\omega,t,\bar{x}(t),\bar{u},\bar{p}(t),\bar{q}(t))(v(t)-\bar{u}(t))\mathrm{d}t\right) \geq 0. \tag{3.4.18}$$

Now, let $(\bar{\alpha}, \bar{\lambda}_{\mathcal{I}}, \bar{\lambda}_E, \bar{\lambda}_I)$ be a generalized Lagrange multiplier at $(\bar{x}, \bar{u})$. By the first relation in (3.4.13) and (3.4.16) we obtain that

$$\bar{\lambda}_{\mathcal{I}} = p(0) + \int_0^{\cdot} p(t)\mathrm{d}t + \int_0^{\cdot} q(t)\mathrm{d}W(t),$$

where $(p, q) \in \mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^{n\times d}$ solves (3.4.17). Therefore, we get $(\bar{\lambda}_{\mathcal{I}})_1 = p$ and $(\bar{\lambda}_{\mathcal{I}})_2 = q$ and $(\bar{\lambda}_{\mathcal{I}})_0 = p(0) = (\bar{\lambda}_{\mathcal{I}})_1(0)$. Thus (3.4.17) implies that $((\bar{\lambda}_{\mathcal{I}})_1, (\bar{\lambda}_{\mathcal{I}})_2)$ satisfies the first and second relations in (3.4.11). Finally, by the second relation in (3.4.13) and expression (3.4.18), we obtain the third relation in (3.4.11) following the same argument as in the proof of [Cadenillas and Karatzas, 1995, Theorem 1.5]. ∎

As a consequence of the above result we obtain the following sufficient condition, under convexity assumptions. The proof is standard, but since it is very short we provide it for the reader's convenience.

Corollary 3.4.1 (Sufficient condition for convex problems) *Suppose that F and G_I are convex and that G and G_E are affine.*

(i) *Let $(\bar{x}, \bar{u}) \in F(\mathcal{SP})$ and suppose that $(\bar{p}, \bar{q}, \bar{\lambda}_E, \bar{\lambda}_I) \in \mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^{n\times d} \times \mathbb{R}^{n_E} \times \mathbb{R}^{n_I}$ is a weak-Pontryagin multiplier at $(\bar{x}, \bar{u})$. Then, the pair $(\bar{x}, \bar{u})$ solves $(\mathcal{SP})$.*

(ii) *The set of weak-Pontryagin multipliers is independent of the solutions of $(\mathcal{SP})$. More precisely, let $(\bar{x}^1, \bar{u}^1)$, $(\bar{x}^2, \bar{u}^2) \in F(\mathcal{SP})$ be two solutions of $(\mathcal{SP})$. Then, $(\bar{p}, \bar{q}, \bar{\lambda}_E, \bar{\lambda}_I)$ is a weak-Pontryagin multiplier at $(\bar{x}^1, \bar{u}^1)$ if and only if it is a weak-Pontryagin multiplier at $(\bar{x}^2, \bar{u}^2)$.*

Proof. By Theorem 3.4.2, $\bar{\lambda}_{\mathcal{I}} \in \mathcal{I}^n$ defined by (3.4.14) is such that $(\bar{\lambda}_{\mathcal{I}}, \bar{\lambda}_E, \bar{\lambda}_I)$ is a Lagrange multiplier. Now, let (x, u) be feasible for $(\mathcal{SP})$, then by the convexity of $\mathcal{L}(\cdot, \cdot, 1, \bar{\lambda}_{\mathcal{I}}, \bar{\lambda}_E, \bar{\lambda}_I)$,

$$\begin{aligned} F(x,u) \geq \mathcal{L}(x, u, 1, \bar{\lambda}_{\mathcal{I}}, \bar{\lambda}_E, \bar{\lambda}_I) \geq \quad & \mathcal{L}(\bar{x}, \bar{u}, 1, \bar{\lambda}_{\mathcal{I}}, \bar{\lambda}_E, \bar{\lambda}_I) + D_x\mathcal{L}(\bar{x}, \bar{u}, 1, \bar{\lambda}_{\mathcal{I}}, \bar{\lambda}_E, \bar{\lambda}_I)(x - \bar{x}) \\ & + D_u\mathcal{L}(\bar{x}, \bar{u}, 1, \bar{\lambda}_{\mathcal{I}}, \bar{\lambda}_E, \bar{\lambda}_I)(u - \bar{u}). \end{aligned}$$

Since $\mathcal{L}(\bar{x}, \bar{u}, 1, \bar{\lambda}_{\mathcal{I}}, \bar{\lambda}_E, \bar{\lambda}_I) = F(\bar{x}, \bar{u})$ assertion (i) follows from (3.4.13). Assertion (ii) is a direct consequence of Theorem 3.4.2 and the fact that for convex problems the set of Lagrange multipliers $\Lambda_L(\bar{x}, \bar{u})$ does not depend on $(\bar{x}, \bar{u})$ (see e.g. [Bonnans and Shapiro, 2000, Theorem 3.6]). ∎

3.5 Some sensitivity results

In this section we take advantage of the Lagrange multiplier interpretation of the adjoint state (p, q) in order to obtain some sensitivity results for the optimal cost when the problem dynamics and final constraints are perturbed. We will first consider general convex problems and linear perturbations of the dynamics. Next, we study in detail the case of Linear Quadratic (LQ) stochastic problems and the mean variance portfolio selection problem, where the perturbations are performed also in the matrices multiplying the state and control variables. We shall study these last two problems separately, since although they belong to a same family, their specific structures mean that we need to employ slightly different arguments and assume different hypotheses. In any case a stability result for the solutions of the parameterized problems is needed and will be a consequence of the following result:

Proposition 3.5.1 *The following assertions hold:*

(i) *Let $x^k \in \mathcal{I}^n$ be a sequence converging weakly to $x \in \mathcal{I}^n$. Then x^k converges weakly to x in $(L^{2,2}_{\mathbb{F}})^n$ and for all $t \in [0,T]$ we have that $x^k(t)$ converges weakly to $x(t)$ in $(L^2_{\mathcal{F}_t})^n$.*

(ii) *Let $x_0^k \in \mathbb{R}^n$, $A^k \in (L^{\infty,\infty}_{\mathbb{F}})^{n\times n}$, $(C^j)^k \in (L^{\infty,\infty}_{\mathbb{F}})^{n\times n}$, $\xi_1^k \in (L^{2,2}_{\mathbb{F}})^n$, $(\xi_2^j)^k \in (L^{2,2}_{\mathbb{F}})^n$ $(j = 1,\dots,d)$. Suppose that $(x_0^k, A^k, (C^j)^k)$ converge strongly to (x_0, A, C^j) and that $(\xi_1^k, (\xi_2^j)^k)$ converge weakly to (ξ_1, ξ_2^j). Then, the solutions x^k of*

$$
\begin{aligned}
\mathrm{d}x^k(t) &= \left[A^k(t)x^k(t) + \xi_1^k(t)\right]\mathrm{d}t + \textstyle\sum_{j=1}^d \left[(C^j)^k(t)x^k(t) + (\xi_2^j)^k(t)\right]\mathrm{d}W^j(t),\\
x^k(0) &= x_0^k,
\end{aligned}
$$

converge weakly in $\mathcal{I}^n$ to the solution x of

$$
\begin{aligned}
\mathrm{d}x(t) &= \left[A(t)x(t) + \xi_1(t)\right]\mathrm{d}t + \textstyle\sum_{j=1}^d \left[C^j(t)x(t) + \xi_2^j(t)\right]\mathrm{d}W^j(t),\\
x(0) &= x_0.
\end{aligned}
\tag{3.5.1}
$$

(iii) *Let $D^k \in (L^{\infty,\infty}_{\mathbb{F}})^{n\times n}$, $(E^j)^k \in (L^{\infty,\infty}_{\mathbb{F}})^{n\times n}$ $(j = 1,\dots,d)$, $\xi_3^k \in (L^{2,2}_{\mathbb{F}})^n$ and $\xi_4^k \in (L^2_{\mathcal{F}_T})^n$. Suppose that $(D^k, (E^j)^k)$ converge strongly to (D, E^j) and (ξ_3^k, ξ_4^k) converge weakly to (ξ_3, ξ_4). Then, the solutions (p^k, q^k) of*

$$
\begin{aligned}
\mathrm{d}p^k(t) &= \left[D^k(t)p^k(t) + \textstyle\sum_{j=1}^d (E^j)^k(t)(q^j)^k(t) + \xi_3^k(t)\right]\mathrm{d}t + q^k(t)\mathrm{d}W(t),\\
p^k(T) &= \xi_4^k.
\end{aligned}
\tag{3.5.2}
$$

converge weakly in $\mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^{n\times d}$ to the solution (p,q) of

$$
\begin{aligned}
\mathrm{d}p(t) &= \left[D(t)p(t) + \textstyle\sum_{j=1}^d E^j(t)q^j(t) + \xi_3(t)\right]\mathrm{d}t + q(t)\mathrm{d}W(t),\\
p(T) &= \xi_4.
\end{aligned}
\tag{3.5.3}
$$

Proof. Assertion (i) follows directly from Lemma 3.2.2 and the fact that $(L^{2,\infty}_{\mathbb{F}})^n$ is continuously embedded in $(L^{2,2}_{\mathbb{F}})^n$ and $(L^2_{\mathcal{F}_t})^n$, for all $t \in [0,T]$. Let us prove assertion (ii). Since $|x_0^k|$, $\|A^k\|_{\infty,\infty}$, $\|(C^j)^k\|_{\infty,\infty}$, $\|(D^j)^k\|_{\infty,\infty}$, $\|\xi_1^k\|_{2,2}$ and $\|(\xi_2^j)^k\|_{2,2}$ are bounded, by the classical proof for the stability of linear SDEs (see e.g. [Yong and Zhou, 1999, Chapter 6, Section 4]), we have that $\|x^k\|_{2,\infty}$ is uniformly bounded in k. Therefore for any subsequence there exists $\hat{x} \in (L^{2,2}_{\mathbb{F}})^n$ such that a further subsequence x^k converges weakly in $(L^{2,2}_{\mathbb{F}})^n$ to $\hat{x}$. Using that $A^k x^k$, $(C^j)^k x^k$ converge weakly in $(L^{2,2}_{\mathbb{F}})^n$ to $A\hat{x}$, $C^j\hat{x}$, respectively, we see that x^k converges weakly in $\mathcal{I}^n$ to

$$
\tilde{x}(\cdot) := x_0 + \int_0^{\cdot} \left[A(t)\hat{x}(t) + \xi_1(t)\right]\mathrm{d}t + \sum_{j=1}^d \int_0^{\cdot} \left[C^j(t)\hat{x}(t) + \xi_2^j(t)\right]\mathrm{d}W(t).
$$

By (i) we have that $\tilde{x} = \hat{x}$ and since (3.5.1) has a unique solution (and so independent of the given subsequence) the result follows. In order to prove (iii), we argue in a similar manner. Note that since (ξ_3^k, ξ_4^k) is bounded in $(L_{\mathbb{F}}^{2,2})^n \times (L_{\mathcal{F}_T}^2)^n$ and $\|D^k\|_{\infty,\infty}$, $\|(E^j)^k\|_{\infty,\infty}$ are bounded, following the lines of the proof [Yong and Zhou, 1999, Chapter 7, Theorem 2.2]) we obtain that $\|p^k\|_{2,\infty} + \sum_{j=1}^d \|(q^j)^k\|_{2,2}$ is uniformly bounded in k. So for any subsequence there exists $(\hat{p}, \hat{q}) \in (L_{\mathbb{F}}^{2,2})^n \times (L_{\mathbb{F}}^{2,2})^{n\times d}$ such that, except for some further subsequence, (p^k, q^k) converge to $(\hat{p}, \hat{q})$ weakly in $(L_{\mathbb{F}}^{2,2})^n \times (L_{\mathbb{F}}^{2,2})^{n\times d}$. Since $D^k p^k$ and $(E^j)^k (q^j)^k$ converge weakly in $(L_{\mathbb{F}}^{2,2})^n$ respectively to $D\hat{p}$, $E^j \hat{q}^j$, we easily obtain that p^k converges weakly in $\mathcal{I}^n$ to

$$\tilde{p}(\cdot) := \tilde{p}(0) + \int_0^{\cdot} \left[D(t)\hat{p}(t) + \sum_{j=1}^d E^j(t)\hat{q}^j(t) + \xi_3(t) \right] \mathrm{d}t + \int_0^{\cdot} \hat{q}(t)\mathrm{d}W(t), \tag{3.5.4}$$

where

$$\tilde{p}(0) := \mathbb{E}\left(\xi_4 - \int_0^T \left[D(t)\hat{p}(t) + \sum_{j=1}^d E^j(t)\hat{q}^j(t) + \xi_3(t) \right] \mathrm{d}t \right).$$

By (i) we obtain that $\tilde{p} = \hat{p}$, and $\hat{p}(T) = \xi_4$ using that $p^k(T) = \xi_4^k$ converges weakly in $(L_{\mathcal{F}_T}^2)^n$ to ξ_4. From this fact and (3.5.4), we have that $(\hat{p}, \hat{q})$ solves (3.5.3). Finally, since the solution of (3.5.3) is unique, the result follows. ■

3.5.1 Convex problems and linear perturbations of the dynamics

Let us define the perturbation space $\mathcal{P}_1 := \mathbb{R}^n \times (L_{\mathbb{F}}^{2,2})^n \times (L_{\mathbb{F}}^{2,2})^{n\times d}$ and let $P := (x_0, \hat{f}, \hat{\sigma}) \in \mathcal{P}_1$. We consider the problem

$$\begin{array}{ll} \inf\limits_{(x,u)\in\mathcal{I}^n\times(L_{\mathbb{F}}^{2,2})^m} & \mathbb{E}\left(\int_0^T \ell(t,\omega,x(t),u(t))\mathrm{d}t + \Phi(\omega,x(T)) \right) \\ \text{s.t.} & \left\{ \begin{array}{rcl} \mathrm{d}x(t) &=& [f(t,\omega,x(t),u(t)) + \hat{f}(t,\omega)]\mathrm{d}t + [\sigma(t,\omega,x(t),u(t)) + \hat{\sigma}(t,\omega)]\mathrm{d}W(t), \\ x(0) &=& x_0, \\ u &\in& \mathcal{U}. \end{array} \right. \end{array} \tag{$P_{1,P}$}$$

We suppose that (ℓ, Φ, f, σ) satisfy Assumptions 11-12 in Section 3.4 and $\mathcal{U}$ is given by (3.4.9). In addition, we will need the following convexity assumption:

Assumption 13 *For almost all $(t,\omega) \in [0,T] \times \Omega$ (respectively $\omega \in \Omega$), the function $\ell(t,\omega,\cdot,\cdot)$ (respectively $\Phi(\omega,\cdot)$) is convex. Moreover, we assume that a.s. in $[0,T] \times \Omega$ the functions $f(t,\omega,\cdot,\cdot)$ and $\sigma(t,\omega,\cdot,\cdot)$ are affine.*

We define *the value function* $v : \mathcal{P}_1 \to \mathbb{R} \cup \{-\infty\}$ for the function that associates to P the optimal cost for problem $(P_{1,P})$. Note that under Assumptions 11-12 the feasible set for $(P_{1,P})$ is not empty, and therefore v is well defined.

In the next theorem we prove that given a nominal parameter P, where problem $(P_{1,P})$ admits at least one solution, and a *random functional perturbation* ΔP acting linearly on the dynamics, the value function v admits directional derivatives that can be expressed in terms of a *unique adjoint state*, which, in view of the convexity of the problem, is independent of the solution of $(P_{1,P})$. The existence of such adjoint state has been proved for the first time in Bismut [1973] and, as pointed out in Remark 3.4.4(i), it also follows from the more general results in Peng [1990]. Due to its simplicity, we provide here a short and direct proof of the existence and uniqueness of such an adjoint state using classical results in abstract optimization theory (see e.g. Maurer and Zowe [1979], Robinson [1976], Zowe and Kurcyusz [1979]) and the identification of Lagrange multipliers and adjoint states in Theorem 3.4.2.

Theorem 3.5.1 *Assume that Assumptions 11-13 hold and that for $P \in \mathcal{P}_1$ problem $(P_{1,P})$ admits at least one solution. Then, there exists $(\bar{p}, \bar{q}) \in \mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^{n\times d}$ such that for every solution $(\bar{x}, \bar{u})$ of $(P_{1,P})$, the pair $(\bar{p}, \bar{q})$ is the unique weak-Pontryagin multiplier associated to $(\bar{x}, \bar{u})$. Moreover, the value function v is continuous at P, Hadamard and Gâteaux directionally differentiable at P and its directional derivative $Dv(P; \cdot) : \mathcal{P}_1 \to \mathbb{R}$ is given by*

$$Dv(P; \Delta P) = \bar{p}(0)^\top \Delta x_0 + \mathbb{E}\left(\int_0^T \bar{p}(t)^\top \Delta f(t) \mathrm{d}t\right) + \mathbb{E}\left(\int_0^T \mathrm{tr}\left[\bar{q}(t)^\top \Delta\sigma(t)\right] \mathrm{d}t\right), \tag{3.5.5}$$

for all $\Delta P = (\Delta x_0, \Delta f, \Delta\sigma) \in \mathcal{P}_1$.

Proof. Let us write the problem $(P_{1,P})$ as

$$\inf_{(x,u)\in\mathcal{I}^n\times(L^{2,2}_{\mathbb{F}})^m} F(x, u) + \chi_{\mathcal{U}}(x, u) \text{ subject to } G(x, u) + P = 0,$$

where $\chi_{\mathcal{U}} : \mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^m \to \mathbb{R} \cup \{+\infty\}$ is the convex, proper, l.s.c. function defined as $\chi_{\mathcal{U}}(x, u) = 0$ if $u \in \mathcal{U}$ and $+\infty$ otherwise and

$$G(x, u)(\cdot) := \int_0^{\cdot} f(t, \omega, x(t), u(t))\mathrm{d}t + \int_0^{\cdot} \sigma(t, \omega, x(t), u(t))\mathrm{d}W(t) - x(\cdot).$$

For every $(x, u) \in \mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^m$ and $v \in (L^{2,2}_{\mathbb{F}})^m$, Lemma 3.4.1 implies that $DG(x, u)(\cdot, v)$ is surjective. Therefore, the following regularity condition is trivially satisfied (see e.g. [Bonnans and Shapiro, 1998, Section 3.2])

$$0 \in \mathrm{int}\left\{G(x, u) + P + DG(x, u)\left(\mathcal{I}^n \times \mathcal{U}\right)\right\}. \tag{3.5.6}$$

Thus, by classical results in convex optimization (see e.g. [Bonnans, 2006, Section 4.3.2, Example 4.51] or [Bonnans and Shapiro, 2000, Section 2.5]) (x, u) is a solution of $(P_{1,P})$ iff there exists $\lambda \in \mathcal{I}^n$ such that

$$(0, 0) \in \partial_{(x,u)}(F(x, u) + \chi_{\mathcal{U}}(x, u)) + DG(x, u)^*\lambda. \tag{3.5.7}$$

Since F is differentiable in $\mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^m$, in particular it is continuous in $\mathcal{I}^n \times \mathcal{U}$, and so (see e.g. [Bonnans and Shapiro, 2000, Remark 2.170]):

$$\partial_{(x,u)}(F(x,u)+\chi_{\mathcal{U}}(x,u)) = \partial_{(x,u)}F(x,u)+\partial_{(x,u)}\chi_{\mathcal{U}}(x,u) \subseteq (D_xF(x,u), D_uF(x,u))+\{0\}\times N_{\mathcal{U}}(u),$$

where $N_{\mathcal{U}}(u) := \{v^* \in (L^{2,2}_{\mathbb{F}})^m \ ; \ \langle v^*, v-u\rangle_{L^2} \leq 0, \ \ \forall \ v \in \mathcal{U}\}$ is the normal cone to $\mathcal{U}$ at u. Using that $DG(x,u)^*\lambda = (D_xG(x,u)^*\lambda, D_uG(x,u)^*\lambda)$, we obtain with (3.5.7)

$$(0,0) \subseteq (D_xF(x,u), D_uF(x,u)) + \{0\} \times N_{\mathcal{U}}(u) + (D_xG(x,u)^*\lambda, D_uG(x,u)^*\lambda),$$

which is equivalent to

$$D_x\mathcal{L}(x,u,\lambda) = 0 \quad \text{and} \quad D_u\mathcal{L}(x,u,\lambda)(v-u) \geq 0 \quad \forall v \in \mathcal{U}. \tag{3.5.8}$$

Therefore, $\lambda \in \Lambda_L(x,u)$ and by Theorem 3.4.2 and the convexity of the associated Hamiltonian we have that $(\bar{p},\bar{q}) := (\lambda_1,\lambda_2)$ is weak-Pontryagin multiplier. Now, let $\lambda^1_{\mathcal{I}}$, $\lambda^2_{\mathcal{I}} \in \Lambda_L(x,u)$. By the first equation in (3.5.8), we get that

$$\left\langle (D_xG(x,u))^* \, (\lambda^1_{\mathcal{I}} - \lambda^2_{\mathcal{I}}), z\right\rangle_{\mathcal{I}} = 0 \ \ \forall \ z \in \mathcal{I}^n, \quad \text{or} \quad (D_xG(x,u))^* \, (\lambda^1_{\mathcal{I}} - \lambda^2_{\mathcal{I}}) = 0.$$

Since, by Lemma 3.4.1, $D_xG(x,u) : \mathcal{I}^n \mapsto \mathcal{I}^n$ is surjective we get that $D_xG(x,u)^*$ is injective, which implies that $\lambda^1_{\mathcal{I}} = \lambda^2_{\mathcal{I}}$ and by Theorem 3.4.2 the weak-Pontryagin multiplier is unique. The independence of the set $\Lambda_L(\cdot)$ over the set of solutions of $(P_{1,P})$ is a consequence of Corollary 3.4.1(ii). Finally, the continuity, the Gâteaux and Hadamard differentiability of v and expression (3.5.5) for $Dv(P;\Delta P)$ are a direct translation of [Rockafellar, 1974, Theorem 17] using the uniqueness of the Lagrange multiplier. ■

In the following remark we underline some simple consequences of Theorem 3.5.1:

Remark 3.5.1 (i) *The gradient of v at P, i.e. the Riesz representative of the bounded linear application $Dv(P;\cdot)$, is given by*

$$\bar{p}(0) + \int_0^{\cdot} \bar{p}(t)\mathrm{d}t + \int_0^{\cdot} \bar{q}(t)\mathrm{d}W(t).$$

(ii) *It is well known* (*see e.g.* [Bonnans and Shapiro, 2000, Section 2.2] *and the references therein*) *that for real-valued functions defined on finite dimensional spaces, Gâteaux differentiability together with Hadamard differentiability imply Fréchet differentiability. Therefore, if the perturbations for problem $(P_{1,P})$ are finite dimensional, then v is Fréchet differentiable at P. This is the case, for example, if the initial condition is perturbed and/or the perturbations of the dynamics have the form $\Delta f(t,\omega) = \xi_0(t,\omega)A_0$, $(\Delta\sigma(t,\omega))^j = \xi_j(t,\omega)A_j$ with ξ_0, $\xi_j \in (L^{\infty,\infty}_{\mathbb{F}})^{n\times n}$ $(j = 1,\ldots,d)$ being fixed, and A_0, $A_j \in \mathbb{R}^n$ being the perturbation parameters. In fact, defining the new states*

$$\mathrm{d}y_0 = 0, \ \ \textit{for } t \in [0,T], \ \ y_0(0) = A_0, \ \ \mathrm{d}y_j = 0, \ \ \textit{for } t \in [0,T], \ \ y_j(0) = A_j \ \ \textit{for } j = 1,\ldots,d,$$

the new dynamical system is affine w.r.t. $(x,(y_0,y_j))$ *and the perturbations are performed over the initial condition. Let us point out that the Fréchet differentiability of the value function under finite-dimensional perturbations in our convex framework can also be deduced using* [Yong and Zhou, 1999, Chapter 5, Corollary 4.5].

(iii) *Suppose that the nominal problem is deterministic (and thus* $\bar{q}=0$*) and only the* $\mathrm{d}W(t)$ *part of the dynamics is perturbed, i.e.* $\Delta x_0 = 0$, $\Delta f \equiv 0$. *Then, by* (3.5.5) *we directly obtain that* $Dv(P;\Delta P)=0$. *This fact was already observed by Loewen in* Loewen [1987] *for finite dimensional perturbations.*

(iv) *A close look at the proof Theorem 3.5.1 shows that even if* $\ell(\omega,t,\cdot,\cdot)$ *and* $\Phi(\omega,\cdot)$ *are not convex, we can apply the abstract optimization results* (*see e.g.* [Bonnans and Shapiro, 2000, Section 3.1]) *in order to derive existence and uniqueness of a Lagrange multiplier at a local solution* $\bar{u}$. *More precisely, using* (3.5.6) *it is possible to show* (*see* [Bonnans and Shapiro, 2000, Lemma 3.7]) *that if* (x,u) *is a solution of problem* $(P_{1,P})$ *then* $(z,v)=(0,0)$ *is a solution of*

$$\inf_{(z,v)\in\mathcal{I}^n\times(L^{2,2}_{\mathbb{F}})^m} DF(x,u)(z,v) \quad \textit{such that} \quad DG(x,u)(z,v)=0, \quad v\in T_{\mathcal{U}}(u), \qquad (LP)$$

(*where* $T_{\mathcal{U}}(u)$, *defined as the closure in* $(L^{2,2}_{\mathbb{F}})^m$ *of* $\bigcup_{\tau>0}\tau^{-1}(\mathcal{U}-u)$, *is the tangent cone to* $\mathcal{U}$ *at* u, *see* [Bonnans and Shapiro, 2000, Proposition 2.55]). *Problem* (LP) *is a convex one and we can proceed exactly as in the proof of Theorem 3.5.1 in order to show the existence and uniqueness of a Lagrange multiplier* λ *at* $(0,0)$. *It is easy to see that* λ *is a Lagrange multiplier at* $(0,0)$ *for problem* (LP) *iff* λ *is a Lagrange multiplier at* (x,u) *for problem* $(P_{1,P})$. *Therefore, by Theorem 3.4.2 this argument provides a simple proof of the existence of weak-Pontryagin multipliers for stochastic problems with non-convex cost and linear dynamics. Let us point out that it is not clear that the general result of* Peng [1990] *for the case of nonlinear dynamics, even in the form of a weak-Pontryagin principle, can be derived with the Lagrange multipliers method. In fact, the main issue is the apparent lack of* C^1 *differentiability of* $G(x,u)$ *in the non-affine case* (*see Remark 3.4.2*) *and the non-convexity of* $\mathcal{U}$ *in* Peng [1990].

We consider now the case of *final state constraints without control constraints*[1]. We propose as parameter set the space $\mathcal{P}_2 := \mathbb{R}^n \times (L^{2,2}_{\mathbb{F}})^n \times (L^{2,2}_{\mathbb{F}})^{n\times d} \times \mathbb{R}^{n_E} \times \mathbb{R}^{n_I}$. Let $P := (x_0,\hat{f},\hat{\sigma},\delta_E,\delta_I)\in\mathcal{P}_2$ and consider the problem

[1] Actually we can handle also control and final state constraints simultaneously under a suitable qualification condition (see [Bonnans and Shapiro, 1998, Section 3.2]). However, for the sake of simplicity we preferred to state the results for both types of constraints separately.

$$\inf_{(x,u)\in\mathcal{I}^n\times(L^{2,2}_{\mathbb{F}})^m} \mathbb{E}\left(\int_0^T \ell(t,\omega,x(t),u(t))\mathrm{d}t + \Phi(\omega,x(T))\right)$$
$$\text{s.t.} \quad \left\{\begin{array}{rcl} \mathrm{d}x(t) & = & [f(t,\omega,x(t),u(t)) + \hat{f}(t,\omega)]\mathrm{d}t \\ & & +[\sigma(t,\omega,x(t),u(t)) + \hat{\sigma}(t,\omega)]\mathrm{d}W(t), \\ x(0) & = & x_0, \\ \mathbb{E}\left(\Phi^i_E(\omega,x(T))\right) & = & -\delta^i \quad \text{for all } i=1,\dots,n_E, \\ \mathbb{E}\left(\Phi^j_I(\omega,x(T))\right) & \leq & -\delta^j \quad \text{for all } j=1,\dots,n_I. \end{array}\right. \tag{$P_{2,P}$}$$

We will assume that:

Assumption 14 *For almost all $\omega\in\Omega$ the functions $\Phi^i_E(\omega,\cdot)$ ($i=1,\dots,n_E$) are affine and $\Phi^j_I(\omega,\cdot)$ ($j=1,\dots,n_I$) are convex.*

The proof of the following result follows the same lines as those in the proof of Theorem 3.5.1 and therefore is omitted. Recall that G is defined in (3.4.2) and G_E, G_I are defined in (3.4.7).

Theorem 3.5.2 *Assume that the Assumptions 11-14 hold and that for $P\in\mathcal{P}_2$ problem ($P_{2,P}$) admits at least one solution $(\bar{x},\bar{u})$. Suppose in addition that the following Slater constraint qualification condition at $(\bar{x},\bar{u})$ holds*

$$\left.\begin{array}{l} \text{(i)}\ \ (DG(\bar{x},\bar{u}),DG_E(\bar{x})):\mathcal{I}^n\times(L^{2,2}_{\mathbb{F}})^m\mapsto\mathcal{I}^n\times\mathbb{R}^{n_E}\ \ \textit{is surjective and} \\ \text{(ii)}\ \ \exists\,(\hat{z},\hat{v})\in\mathcal{I}^n\times(L^{2,2}_{\mathbb{F}})^m\ \ ;\ G(\hat{z},\hat{v})=0,\ \ G_E(\hat{z},\hat{v})=0,\ \ G^j_I(\hat{z})<0\ \forall\,j=1,\dots,,n_I. \end{array}\right\} \tag{S}$$

Then, the set of weak-Pontryagin multipliers $\Lambda_{wP}(\bar{x},\bar{u})\subset\mathcal{I}^n\times\mathbb{R}^{n_E+n_I}$ at any solution $(\bar{x},\bar{u})$ is a non-empty, weakly compact set, which is independent of the solution $(\bar{x},\bar{u})$. Moreover, the value function v is continuous at P, Hadamard directionally differentiable at P and its directional derivative $Dv(P;\cdot):\mathcal{P}_2\to\mathbb{R}$ is given by

$$\begin{aligned} Dv(P;\Delta P) = \max_{(p,q,\lambda_E,\lambda_I)\in\Lambda_{wP}(\bar{x},\bar{u})} & \left\{p(0)^\top\Delta x_0 + \mathbb{E}\left(\int_0^T p(t)^\top\Delta f(t)\mathrm{d}t\right)\right. \\ & \left. + \mathbb{E}\left(\int_0^T \mathrm{tr}\left[q(t)^\top\Delta\sigma(t)\right]\mathrm{d}t\right) + \lambda_E^\top\Delta\delta_E + \lambda_I^\top\Delta\delta_I\right\}, \end{aligned}$$

for all $\Delta P=(\Delta x_0,\Delta f,\Delta\sigma,\Delta\delta_E,\Delta\delta_I)\in\mathcal{P}_2$.

Remark 3.5.2 (i) *Note that if no inequality constraints are present (which can be written as $n_I=0$), the qualification condition for ($P_{2,P}$) is given by (S)(i). In this case, as in Theorem 3.5.1, we get the uniqueness of the multiplier and thus v is also Gâteaux differentiable at P.*

(ii) *Since (G,G_E) is affine and G^j_I ($j=1,\dots,n_I$) are convex, we have that the Slater condition (S) is equivalent to the following Mangasarian-Fromovitz condition*

$$\left.\begin{array}{l} \text{(a)}\ \ (DG(\bar{x},\bar{u}),DG_E(\bar{x})):\mathcal{I}^n\times(L^{2,2}_{\mathbb{F}})^m\mapsto\mathcal{I}^n\times\mathbb{R}^{n_E}\ \ \textit{is surjective and} \\ \text{(b)}\ \ \exists\,(\hat{z},\hat{v})\in\mathrm{Ker}DG(\bar{x},\bar{u})\cap\mathrm{Ker}DG_E(\bar{x})\ \ \textit{such that}\ DG^j_I(\bar{x})\hat{z}<0\ \forall\,j=1,\dots,,n_I. \end{array}\right\} \tag{MF}$$

Condition (MF) *has been stated in the literature* (*see e.g.* Bonnans and Silva [2012]) *for the reduced optimal control problem* $(\mathcal{SP}')$. *More precisely, for* $v \in (L^{2,2}_{\mathbb{F}})^m$ *let* $z[v] \in \mathcal{I}^n$ *be defined by the equation* $DG(\bar{x}, \bar{u})(z, v) = 0$. *Since this is a standard linear SDE in the variable* z, *under our assumptions, we get that* $z[v]$ *is well defined. We check then that* (MF) *is equivalent to*

$$\left.\begin{array}{ll} (\mathrm{a}') & v \in (L^{2,2}_{\mathbb{F}})^m \to DG_E(\bar{x})z[v] \in \mathbb{R}^{n_E} \ \ \textit{is surjective and} \\ (\mathrm{b}') & \exists\, \hat{v} \in (L^{2,2}_{\mathbb{F}})^m \ \ \textit{such that} \ \ DG_E(\bar{x})z[v] = 0 \ \ \textit{and} \ DG_I^j(\bar{x})z[v] < 0 \ \forall\, j = 1, \ldots, , n_I. \end{array}\right\}$$

3.5.2 Multiplicative perturbations in the Linear Quadratic framework

In this part we adopt the framework of *unconstrained* Linear Quadratic (LQ) stochastic control problems with random coefficients (see e.g Bismut [1976b], Chen and Yong [2001], Tang [2003], Yong and Zhou [1999] and the references therein). More precisely, let us consider the problem

$$\begin{array}{l} \displaystyle\inf_{(x,u)\in\mathcal{I}^n\times(L^{2,2}_{\mathbb{F}})^m} F(x,u) := \frac{1}{2}\mathbb{E}\left(\int_0^T \left[x(t)^\top Q(t)x(t) + u(t)^\top N(t)u(t)\right] \mathrm{d}t + x(T)^\top Mx(T)\right) \\ \text{s.t.} \ \left\{ \begin{array}{l} \mathrm{d}x(t) = [A(t)x(t) + B(t)u(t) + e(t)]\mathrm{d}t + \displaystyle\sum_{j=1}^d [C^j(t)x(t) + D^j(t)u(t) + f^j(t)]\mathrm{d}W^j(t), \\ x(0) \ = x_0. \end{array} \right. \end{array} \tag{$P_{3,P}$}$$

We shall view $P = (x_0, A, B, C^j, D^j, e, f^j)$ $(j = 1, \ldots, d)$ as parameters for the problem $(P_{3,P})$. Thus, we consider as parameter space

$$\mathcal{P}_3 = \mathbb{R}^n \times (L^{\infty,\infty}_{\mathbb{F}})^{n\times n} \times (L^{\infty,\infty}_{\mathbb{F}})^{n\times m} \times (L^{\infty,\infty}_{\mathbb{F}})^{(n\times n)\times d} \times (L^{\infty,\infty}_{\mathbb{F}})^{(n\times m)\times d} \times (L^{2,2}_{\mathbb{F}})^n \times (L^{2,2}_{\mathbb{F}})^{n\times d}.$$

It is well known (see [Bismut, 1976b, Theorem 2.1]) that given $P \in \mathcal{P}_3$ and $u \in (L^{2,2}_{\mathbb{F}})^m$ the linear SDE in $(P_{3,P})$ admits a unique solution in $\mathcal{I}^n$. We will also need the following result:

Lemma 3.5.1 *The constraint function* $G : \mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^m \times \mathcal{P}_3 \mapsto \mathcal{I}^n$ *defined by:*

$$\begin{array}{rcl} G(x,u,P) & := & -x(\cdot) + x_0 + \int_0^{\cdot} [A(t)x(t) + B(t)u(t) + e(t)]\mathrm{d}t \\ & & + \int_0^{\cdot} \displaystyle\sum_{j=1}^d [C^j(t)x(t) + D^j(t)u(t) + f^j(t)]\mathrm{d}W^j(t), \end{array}$$

is continuously Fréchet differentiable. Furthermore, $D_{(x,u)}G(x,u,P)$ *is onto.*

Proof. That G is well defined is a simple application of Lemma 3.2.2. Following the lines of the proof of Lemma 3.4.1 we have that G is Gâteaux differentiable at any $(x, u, P) \in \mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^m \times \mathcal{P}_3$ and for every $(x', u') \in \mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^m$ and $P' = (x_0', A', B', \{(C^j)'\}, \{(D^j)'\}, e', \{(f^j)'\}) \in \mathcal{P}_3$ we have that

$$\begin{array}{rcl} DG(x,u,P)(x',u',P') & = & \int_0^{\cdot} [Ax' + Bu' + e']\mathrm{d}t + \int_0^{\cdot} \displaystyle\sum_{j=1}^d [C^jx' + D^ju' + (f')^j]\mathrm{d}W^j \\ & & + \int_0^{\cdot} [A'x + B'u]\mathrm{d}t + \int_0^{\cdot} \displaystyle\sum_{j=1}^d [(C^j)'X + (D^j)'u]\mathrm{d}W^j + x_0' - x'(\cdot). \end{array}$$

Thus, for every (x_1, u_1), $(x_2, u_2) \in \mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^m$ and P_1, $P_2 \in \mathcal{P}_3$ we have that

$$\|DG(x_1, u_1, P_1)(x', u', P') - DG(x_2, u_2, P_2)(x', u', P')\|^2_{\mathcal{I}}$$

is given by

$$\begin{aligned} &\mathbb{E}\left(\int_0^T |(A_1 - A_2)x' + (B_1 - B_2)u' + A'(x_1 - x_2) + B'(u_1 - u_2)|^2 \,\mathrm{d}t\right) + \\ &\mathbb{E}\left(\sum_{j=1}^d \int_0^T \left|(C_1^j - C_2^j)x' + (D_1^j - D_2^j)u' + (C^j)'(x_1 - x_2) + (D^j)'(u_1 - u_2)\right|^2 \mathrm{d}t\right). \end{aligned}$$

Therefore, if $\|P'\| = 1$, we find that $\|DG(x_1, u_1, P_1)P' - DG(x_2, u_2, P_2)P'\|^2_{\mathcal{I}}$ is bounded by

$$c\left(\|x_1 - x_2\|^2_{\mathcal{I}} + \|u_1 - u_2\|^2_2 + \|A_1 - A_2\|^2_\infty + \|B_1 - B_2\|^2_\infty + \sum_{j=1}^d \left[\|C_1^j - C_2^j\|^2_\infty + \|D_1^j - D_2^j\|^2_\infty\right]\right),$$

for some $c > 0$, where we used Lemma 3.2.2 to make $\|\cdot\|_{\mathcal{I}}$ appear. Thus, G is Gâteaux differentiable with a continuous directional derivative, and so G is indeed Fréchet continuously differentiable. The surjectivity of $D_{(x,u)}G(x, u, P)$ follows from Lemma 3.4.1. ■

We make the following convexity assumption:

Assumption 15 *The matrix processes $Q : [0, T] \times \Omega \mapsto \mathbb{R}^{n\times n}$, $N : [0, T] \times \Omega \mapsto \mathbb{R}^{m\times m}$ are essentially bounded and progressively measurable, whereas the matrix $M : \Omega \mapsto \mathbb{R}^{n\times n}$ is essentially bounded and $\mathcal{F}_T$-measurable. In addition Q, N and M are a.s. non-negative symmetric matrices and further there exists $\delta > 0$ such that $N \succeq \delta I$.*

By [Bismut, 1976b, Theorem 3.1] we have that under Assumption 15 problem $(P_{3,P})$ admits a unique solution $(x[P], u[P])$. Moreover, by [Bismut, 1976b, Theorem 3.2] (or Theorem 3.5.1) we obtain the existence of a unique weak-Pontryagin multiplier $(p[P], q[P]) \in \mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^{n\times d}$ such that

$$\begin{array}{rcl} \mathrm{d}x(t) &=& [A(t)x(t) + B(t)u(t) + e(t)]\mathrm{d}t + \sum\limits_{j=1}^d [C^j(t)x(t) + D^j(t)u(t) + f^j(t)]\mathrm{d}W^j(t), \\ u(t) &=& -N(t)^{-1}\left[B(t)^\top p(t) + \sum_{j=1}^d D^j(t)^\top q^j(t)\right], \\ \mathrm{d}p(t) &=& -[A(t)^\top p(t) + \sum_{j=1}^d C^j(t)^\top q^j(t) + Q(t)x(t)]\mathrm{d}t + \sum_{j=1}^d q^j(t)\mathrm{d}W^j(t), \\ x(0) = x_0 &,& p(T) = Mx(T), \end{array} \tag{3.5.9}$$

where we have omitted the dependence on P in order to simplify the notation. We want to obtain now an energy estimate for $(x[P], u[P], p[P], q[P])$ in terms of P, in the spirit of [Tang, 2003, Theorem 2.2]. Because we need to keep track of the constants that will appear (since they depend on model parameters, which we shall later vary) we prove the following Lemma:

Lemma 3.5.2 *Under Assumption 15 there exists a continuous function $\beta : \mathcal{P}_3 \to \mathbb{R}$*

such that

$$\|x[P]\|_{\mathcal{I}}^2 + \|u[P]\|_{2,2}^2 + \|p[P]\|_{\mathcal{I}}^2 + \sum_{j=1}^{d} \|q^j[P]\|_{2,2}^2 \leq \beta(P).$$

Proof. For notational convenience we will omit the dependence on P of the vector $(x[P], u[P], p[P], q[P])$. A close look at the classical proof for the stability of solutions to linear SDEs (see e.g. [Yong and Zhou, 1999, Chapter 6, Section 4]) and of linear BSDEs (see e.g. [Yong and Zhou, 1999, Chapter 7, Theorem 2.2]) gives that

$$\begin{array}{rcl} \|x\|_{2,\infty}^2 & \leq & \kappa_0(P)\left(\|u\|_{2,2}^2 + |x_0|^2 + \|e\|_{2,2}^2 + \sum_{j=1}^d \|f^j\|_{2,2}^2\right), \\ \|p\|_{2,\infty}^2 + \sum_{j=1}^d \|q^j\|_{2,2}^2 & \leq & \kappa_1(P)\mathbb{E}\left(|M(T)x(T)|^2 + \int_0^T |Q(t)x(t)|^2 \mathrm{d}t\right), \end{array} \tag{3.5.10}$$

where

$$\kappa_0 = \kappa_0\left(\|A\|_{\infty,\infty}, \|B\|_{\infty,\infty}, , \sum \|C^j\|_{\infty,\infty}, \sum \|D^j\|_{\infty,\infty}\right), \quad \kappa_1 = \kappa_1\left(\|A\|_{\infty,\infty}, \sum \|C^j\|_{\infty,\infty}\right),$$

are continuous functions. Recall that for a symmetric non-negative matrix $L \in \mathbb{R}^{n\times n}$ one has that $k_L L \succeq L^2$ for k_L equals the largest eigenvalue of L. It is easy to check that $k_L \leq n \max_{i,j\in\{1,\dots,n\}} |L^{ij}|$. Applying this we see that

$$\begin{array}{rcl} \int_0^T |Q(t)x(t)|^2 \mathrm{d}t & \leq & c\int_0^T x(t)^\top Q(t)x(t)\mathrm{d}t, \\ |M(T)x(T)|^2 & \leq & cx(T)^\top M(T)x(T), \end{array} \tag{3.5.11}$$

where $c = n\max\{\|Q\|_{\infty,\infty}, \|M\|_\infty\}$. Now, combining Lemma 3.3.2 and (3.5.9), we get

$$\mathbb{E}\left(x(T)^\top M(T)x(T) + \int_0^T \left[x^\top Qx + u^\top Nu\right]\mathrm{d}t\right) = p(0)^\top x_0 + \mathbb{E}\left(\int_0^T \left[p^\top e + \sum_{j=1}^d (q^j)^\top f^j\right]\mathrm{d}t\right). \tag{3.5.12}$$

Therefore, by the second inequality in (3.5.10), (3.5.11) and (3.5.12) we have that

$$\|p\|_{2,\infty}^2 + \sum_{j=1}^d \|q^j\|_{2,2}^2 \leq c\kappa_1\left\{|p(0)||x_0| + \mathbb{E}\left(\int_0^T \left|p^\top e + \sum_{j=1}^d (q^j)^\top f^j\right|\mathrm{d}t\right)\right\}.$$

Using now the inequality $2ab \leq a^2 + b^2$ for all a, $b \in \mathbb{R}$, we get that

$$\|p\|_{2,\infty}^2 + \sum_{j=1}^d \|q^j\|_{2,2}^2 \leq \kappa_2\mathbb{E}\left(\left[\int_0^T \left(\frac{|x_0|}{T} + |e|\right)\mathrm{d}t\right]^2 + \int_0^T \sum_{j=1}^d |f^j|^2 \mathrm{d}t\right), \tag{3.5.13}$$

where κ_2 depends continuously on c and κ_1 only, and so the r.h.s. is clearly a continuous function of the model parameters. On the other hand, by (3.5.11) we have that

$$\delta\|u\|_{2,2}^2 \leq p(0)^\top x_0 + \mathbb{E}\left(\int_0^T \left|p^\top e + \sum_{j=1}^d (q^j)^\top f^j\right|\mathrm{d}t\right).$$

Using (3.5.13) we obtain that $\|u\|_{2,2}^2$ is bounded by a continuous function of P. Therefore, from the first equation in (3.5.10) we get that $\|x\|_{2,\infty}^2$ is bounded by a continuous function of P. Thus, noting that

$$
\begin{aligned}
&x_1[P] = Ax[P] + Bu[P] + e \quad \text{and} \quad x_2^j[P] = C^j(t)x[P](t) + D^j(t)u[P](t) + f^j(t), \\
&p_1[P] = -[A(t)^\top p[P](t) + \textstyle\sum_{j=1}^d C^j(t)^\top q[P]^j(t) + Q(t)x[P](t)] \quad \text{and} \quad p_2^j[P] = q[P]^j, \\
&p_0[P] = \mathbb{E}\left(Mx[P](T) - \textstyle\int_0^T p_1[P](t)\mathrm{d}t\right),
\end{aligned}
$$

we obtain that $\|x[P]\|_{\mathcal{I}}^2 + \|p[P]\|_{\mathcal{I}}^2$ is bounded by a continuous function of P. The result follows. ■

We prove now a crucial stability result for the solutions of $(P_{3,P})$ in terms of P. More precisely, let P^k and $P \in \mathcal{P}_3$ be such that $P^k \to P$ as $k \to \infty$. We have the following stability result for $(x^k, u^k, p^k, q^k) := (x[P^k], u[P^k], p[P^k], q[P^k])$.

Proposition 3.5.2 *Suppose that Assumption 15 holds true. Then, as $k \uparrow \infty$, we have that $v(P^k) \to v(P)$ and (x^k, u^k, p^k, q^k) converges strongly in $\mathcal{I}^n \times (L_{\mathbb{F}}^{2,2})^m \times \mathcal{I}^n \times (L_{\mathbb{F}}^{2,2})^{n\times d}$ to $(\bar{x}, \bar{u}, \bar{p}, \bar{q}) := (x[P], u[P], p[P], q[P])$.*

Proof. Let us first prove the convergence of the value functions. Define $\hat{x}^k$ as the solution of the following SDE:

$$
\begin{aligned}
\mathrm{d}\hat{x}^k(t) &= [A^k(t)\hat{x}^k(t) + B^k(t)\bar{u}(t) + e^k(t)]\mathrm{d}t \\
&\quad + \sum_{j=1}^d [(C^j)^k(t)\hat{x}^k(t) + (D^j)^k(t)\bar{u}(t) + (f^j)^k(t)]\mathrm{d}W^j(t), \\
\hat{x}^k(0) &= x_0^k.
\end{aligned}
$$

By definition, $(\hat{x}^k, \bar{u}) \in F(P_{3,P^k})$ and by the first estimate in (3.5.10) we have $\hat{x}^k$ is bounded in $(L_{\mathbb{F}}^{2,\infty})^n$, uniformly in k. Now, $\hat{z}^k := \hat{x}^k - \bar{x} \in \mathcal{I}^n$ satisfies

$$
\begin{aligned}
\mathrm{d}\hat{z}^k(t) &= [A(t)\hat{z}^k(t) + \delta^k A\hat{x}^k + \delta^k B(t)\bar{u}(t) + \delta^k e(t)]\mathrm{d}t \\
&\quad + \sum_{j=1}^d [C^j(t)\hat{z}^k(t) + \delta^k C^j(t)\hat{x}^k + \delta^k D^j(t)\bar{u}(t) + \delta^k f^j(t)]\mathrm{d}W^j(t) \\
\hat{z}^k(0) &= \delta^k x_0,
\end{aligned}
$$

where $\delta^k A := A^k - A$, $\delta^k B := B^k - B$ and $\delta^k e := e^k - e$ with an analogous definition for $\delta^k x_0, \delta^k C^j$, $\delta^k D^j$, $\delta^k f^j$. By the convergence $P^k \to P$, the boundedness of $\hat{x}^k$ in $(L_{\mathbb{F}}^{2,\infty})^n$ and classical bounds for linear SDEs (see e.g. [Yong and Zhou, 1999, Chapter 6, Section 4]), we get that $\hat{z}^k \to 0$ in $(L_{\mathbb{F}}^{2,\infty})^n$. This implies that $|F(\hat{x}^k, \bar{u}) - F(\bar{x}, \bar{u})|$ tends to zero as $k \uparrow \infty$. Therefore, we get

$$
v(P^k) \leq F(\hat{x}^k, \bar{u}) = F(\bar{x}, \bar{u}) + o(1) = v(P) + o(1),
$$

which implies that $\limsup_{k\uparrow\infty}[v(P^k)-v(P)]\le 0$. Analogously, if $\tilde{x}^k$ is the solution of

$$\begin{array}{rcl} \mathrm{d}\tilde{x}^k(t) &=& [A(t)\tilde{x}^k(t)+B(t)u^k(t)+e(t)]\mathrm{d}t \\ && +\sum_{j=1}^{d}[C^j(t)\tilde{x}^k(t)+D^j(t)u^k(t)+f^j(t)]\mathrm{d}W^j(t), \\ \tilde{x}^k(0) &=& x_0, \end{array}$$

we have that $(\tilde{x}^k,u^k)\in F(P_{3,P})$. In addition, $\tilde{z}^k:=x^k-\tilde{x}^k$ satisfies

$$\begin{array}{rcl} \mathrm{d}\tilde{z}^k(t) &=& [A^k(t)\tilde{z}^k(t)+\delta^kA\tilde{x}^k+\delta^kB(t)u^k(t)+\delta^ke(t)]\mathrm{d}t \\ && +\sum_{j=1}^{d}[C^j(t)\tilde{z}^k(t)+\delta^kC^j(t)\tilde{x}^k+\delta^kD^j(t)u^k(t)+\delta^kf^j(t)]\mathrm{d}W^j(t), \\ \tilde{z}^k(0) &=& \delta^kx_0. \end{array}$$

By Lemma 3.5.2 we see that u^k is bounded in $(L^{2,2})^m$. So as before since $P_k\to P$ we get that $\tilde{x}^k$ is bounded in $(L^{2,\infty}_{\mathbb{F}})^n$, and similarly obtain that $\tilde{z}^k\to 0$ in $(L^{2,\infty}_{\mathbb{F}})^n$ and so $|F(\tilde{x}^k,u^k)-F(x^k,u^k)|\to 0$. Thus, we obtain

$$v(P)\le F(\tilde{x}^k,u^k)=F(x^k,u^k)+o(1)=v(P^k)+o(1),$$

which implies that $\liminf_{k\uparrow\infty}[v(P^k)-v(P)]\ge 0$, proving the convergence of the value functions. On the other hand, since P^k converges to P, Lemma 3.5.2 implies the existence of $(\hat{x},\hat{u},\hat{p},\hat{q})$ such that, up to some subsequence, (x^k,u^k,p^k,q^k) converges weakly in $\mathcal{I}^n\times(L^{2,2}_{\mathbb{F}})^m\times\mathcal{I}^n\times(L^{2,2}_{\mathbb{F}})^{n\times d}$ to $(\hat{x},\hat{u},\hat{p},\hat{q})$. By Proposition 3.5.1, we easily get that $(\hat{x},\hat{u},\hat{p},\hat{q})$ satisfies (3.5.9). By Corollary 3.4.1, we have that $(\hat{x},\hat{u})$ is a solution of $(P_{3,P})$, which by uniqueness implies that $(\hat{x},\hat{u})=(\bar{x},\bar{u})$ and so $(\hat{p},\hat{q})=(\bar{p},\bar{q})$. On the other hand, using the elementary fact that for every sequences a_k, b_k of real numbers such that $a^k+b^k\to a+b$ and $a\le\liminf a^k$, $b\le\liminf b^k$ we have that $a^k\to a$ and $b^k\to b$, we get, by the lower semicontinuity of the three terms appearing in F, that $\mathbb{E}\left[\int_0^T(u^k)^\top Nu^k\right]\to\mathbb{E}\left[\int_0^T u^\top Nu\right]$ and so by expanding $\mathbb{E}\left[\int_0^T(u^k-u)^\top N(u^k-u)\right]$ and Assumption 15 we conclude that $\|u^k\|_{2,2}\to\|u\|_{2.2}$. Therefore $u^k\to\bar{u}$ strongly in $(L^{2,2}_{\mathbb{F}})^m$. Setting $z^k:=x^k-\bar{x}$ and $v^k=u^k-\bar{u}$, we have

$$\begin{array}{rcl} \mathrm{d}z^k(t) &=& [A(t)z^k(t)+\delta^kAx^k+B(t)v^k+\delta^kB(t)u^k(t)+\delta^ke(t)]\mathrm{d}t \\ && +\sum_{j=1}^{d}[C^j(t)z^k(t)+\delta^kC^j(t)x^k+D^j(t)v^k+\delta^kD^j(t)u^k(t)+\delta^kf^j(t)]\mathrm{d}W^j(t), \\ z^k(0) &=& \delta^kx_0. \end{array}$$

Since $v^k\to 0$ in $(L^{2,2}_{\mathbb{F}})^m$, using the first estimate of (3.5.10) and the fact that (x^k,u^k) is bounded in $\mathcal{I}^n\times(L^{2,2}_{\mathbb{F}})^m$, we obtain that $x^k\to x$ strongly in $(L^{2,\infty}_{\mathbb{F}})^n$ and consequently, passing to the $(L^{2,2}_{\mathbb{F}})^n$ limit in x_1^k and x_2^k, also in $\mathcal{I}^n$. Finally, setting $\hat{p}^k:=p^k-\bar{p}$ and

$\hat{q}^k := q^k - \bar{q}$, we have that

$$\begin{array}{rl} \mathrm{d}\hat{p}^k(t) = & -\Big[A(t)^\top \hat{p}^k(t) + \delta^k A(t) p^k(t) + \sum_{j=1}^d [C^j(t)^\top (\hat{q})^j(t) + \delta^k C^j(t)^\top (q^j)^k(t)] \\ & +Q(t) z^k(t)\Big]\,\mathrm{d}t + \sum_{j=1}^d (\hat{q}^j)^k(t)\mathrm{d}W^j(t), \\ \hat{p}^k(T) = & M z^k(T). \end{array}$$

Then, applying the classical estimates for linear BSDEs (see e.g. [Yong and Zhou, 1999, Chapter 7, Theorem 2.2]) and using that $z^k(T) \to 0$ strongly in $(L^2_{\mathcal{F}_T})^n$, and that (p^k, q^k) remain bounded in $\mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^{n\times d}$, we get that $(\hat{p}^k, \hat{q}^k) \to (0,0)$ strongly in $(L^{2,2}_{\mathbb{F}})^n \times (L^{2,2}_{\mathbb{F}})^{n\times d}$. By passing to the limit in $\hat{p}_1^k$ and $\hat{p}_2^k$ we obtain the desired result. ■

Define now the value function $v : \mathcal{P}_3 \mapsto \mathbb{R}$ of $(P_{3,P})$ as a function of the parameters. Note that, under Assumption 15, v is well defined. With the previous proposition, we can prove the following sensitivity result:

Theorem 3.5.3 *Suppose that Assumption 15 holds. Then, v is of class C^1. Moreover, at any $P = (x_0, A, B, \{C^j\}, \{D^j\}, e, \{f^j\}) \in \mathcal{P}_3$ the directional derivative is given by*

$$\begin{array}{rcl} Dv(P;\Delta P) & = & \bar{p}(0)\Delta x_0 + \mathbb{E}\left(\int_0^T \bar{p}(t)^\top \left[\Delta A(t)\bar{x}(t) + \Delta B(t)\bar{u}(t) + \Delta e(t)\right]\mathrm{d}t\right) \\ & & +\mathbb{E}\left(\int_0^T \sum_{j=1}^d \bar{q}^j(t)^\top \left[\Delta C^j(t)\bar{x}(t) + \Delta D^j(t)\bar{u}(t) + \Delta f^j(t)\right]\mathrm{d}t\right), \end{array} \tag{3.5.14}$$

where $\Delta P := (\Delta x_0, \Delta A, \Delta B, \{\Delta C^j\}, \{\Delta D^j\}, \Delta e, \{\Delta f^j\})$ and the nominal solution system is $(\bar{x}, \bar{u}, \bar{p}, \bar{q}) = (x[P], u[P], p[P], q[P])$.

Proof. The Hadamard differentiability property for v and expression (3.5.14) follow from the surjectivity result in Lemma 3.5.1, the strong stability of the solutions proved in Proposition 3.5.2, the identification of the Lagrange multipliers with the weak-Pontryagin multipliers proved in Theorem 3.4.2 and [Bonnans and Shapiro, 2000, Theorem 4.24], dealing with sensitivity results for the optimal value in optimization problems in Banach spaces. Moreover, using again Proposition 3.5.2 and expression (3.5.14) we easily check that $Dv(\cdot) : \mathcal{P}_3 \to L(\mathcal{P}_3, \mathbb{R})$ is continuous, which implies the C^1 property. ■

Remark 3.5.3 (i) *Note that if the nominal problem is deterministic then*

$$Dv(P;\Delta P) = \bar{p}(0)\Delta x_0 + \int_0^T \bar{p}(t)^\top \left[\mathbb{E}\left(\Delta A(t)\right)\bar{x}(t) + \mathbb{E}\left(\Delta B(t)\right)\bar{u}(t) + \mathbb{E}[\Delta e(t)]\right]\mathrm{d}t$$

Therefore, the first order term of $v(P+\Delta P) - v(P)$ can be computed with the help of a deterministic differential Riccati equation. This could be useful in practice, since it provides a first order approximation for the value $v(P+\Delta P)$ of the stochastic LQ problem, whose solution is typically characterized in terms of Riccati backward stochastic differential equations, which are more difficult to solve than their deterministic counterpart.

(ii) *It could be interesting to study the extension of the above result for the case of in-*

definite control weight costs, i.e. when N is not necessarily definite positive (*see* Chen et al. [1998], [Yong and Zhou, 1999, Chapter 6] *and references therein*).

3.5.3 Mean-Variance Portfolio Selection

Suppose that a market consists of $d+1$ assets $S^0, S^1, \ldots, S^d$ whose prices are defined by

$$\begin{aligned} \mathrm{d}S^0(t) &= rS^0(t), \text{ for } t \in [0,T], \;\; S^0(0) = 1, \\ dS(t) &= \mathrm{diag}(S(t))\mu(t)\mathrm{d}t + \mathrm{diag}(S(t))\sigma(t)\mathrm{d}W(t) \;\text{ for }\; t \in [0,T],\; S(0) = S_0 \in \mathbb{R}^d, \end{aligned} \tag{3.5.15}$$

where $S := (S^1, \ldots, S^d)$ and for $a \in \mathbb{R}^d$ the matrix $\mathrm{diag}(a) \in \mathbb{R}^{d\times d}$ is defined as $\mathrm{diag}(a)^{ij} = \delta^{ij}ai$ for all $i,j \in \{1,\ldots,d\}$ (δ_{ij} is the Kronecker symbol). The precise properties on the processes $r \in L^\infty([0,T];\mathbb{R})$, $\mu \in (L_{\mathbb{F}}^{\infty,\infty})^d$ and $\sigma \in (L_{\mathbb{F}}^{\infty,\infty})^{d\times d}$ shall be given shortly and will imply that the financial market is arbitrage-free and complete (see e.g. [Karatzas and Shreve, 1998, Chapter 1,Theorem 4.2 and 6.6]).

Given an initial wealth $x \in \mathbb{R}$ and a *self-financing portfolio* $\pi \in (L_{\mathbb{F}}^{2,2})^d$ measured in units of wealth, the associated *wealth process* X is defined through the SDE:

$$\begin{aligned} \mathrm{d}X(t) &= \{r(t)X(t) + \pi(t)^\top(\mu(t) - r(t)\mathbf{1})\}\mathrm{d}t + \pi(t)^\top\sigma(t)\mathrm{d}W(t) \;\text{ for all }\; t \in [0,T], \\ X(0) &= x. \end{aligned} \tag{3.5.16}$$

where $\mathbf{1}$ denotes the vector of ones in $\mathbb{R}^d$. For $A \in \mathbb{R}$ we consider the problem (see e.g. Duffie and Richardson [1991], Li and Zhou [2000], Framstad et al. [2004]):

$$\inf_{(X,\pi)\in\mathcal{I}^1\times(L_{\mathbb{F}}^{2,2})^d} \mathbb{E}\left([X-A]^2\right), \;\text{ such that (3.5.16) is verified and }\; \mathbb{E}\left(X(T)\right) = A. \tag{MVP}$$

We then see that the aim is to minimize the risk (variance) subject to a guaranteed mean-return at the final time T.

We intend to compute the sensitivities of this problem with respect to its parameters. Let us define as *parameter space* $\mathcal{P}_4 := \mathbb{R} \times L^\infty([0,T]) \times \mathbb{R} \times (L_{\mathbb{F}}^{\infty,\infty})^d \times (L_{\mathbb{F}}^{\infty,\infty})^{d\times d}$. We will further say that $P = (x, r, A, \mu, \sigma)$ belongs to $\hat{\mathcal{P}}_4$ if $P \in \mathcal{P}_4$, $\sigma\sigma^\top \succeq \delta I_{d\times d}$ for some $\delta > 0$, and

$$\sum_{i=1}^d \left| \mathbb{E}\left(\int_0^T [\mu_i(t) - r(t)]\mathrm{d}t \right) \right| > 0. \tag{3.5.17}$$

Note that $\hat{\mathcal{P}}_4$ is an open subset of $\mathcal{P}_4$. Let us call $v(P) :=$ *value of* (MVP), the corresponding optimal value function (as a function of the model parameters). On a first step we prove some estimates relating the norms of the portfolio and wealth. As in the LQ-case, we compute the constants rather explicitly to show that they will not explode when we vary the model parameters.

Lemma 3.5.3 *If $P = (x, r, A, \mu, \sigma) \in \hat{\mathcal{P}}_4$ and X satisfies* (3.5.16)*, then*

$$\begin{aligned} \|\pi\|_{2,2}^2 &\leq \frac{2}{\delta}\mathbb{E}\left[X(T)^2\right]\left(1 + 2T\left(\|r\|_\infty + \|\sigma^{-1}\{\mu - r\mathbf{1}\}\|_\infty^2\right)e^{2\left(\|r\|_\infty + \|\sigma^{-1}\{\mu - r\mathbf{1}\}\|_\infty^2\right)T}\right), \\ \|X\|_{2,2}^2 &\leq T\mathbb{E}\left(|X(T)|^2\right)e^{2\left(\|r\|_\infty + \|\sigma^{-1}\{\mu - r\mathbf{1}\}\|_\infty^2\right)T}. \end{aligned} \tag{3.5.18}$$

Proof. By classical results on SDEs (e.g. [Bismut, 1976b, Theorem 2.1]) we have that $X \in L_{\mathbb{F}}^{2,\infty}$. Let us set $Z = \sigma^\top \pi$. We have that

$$X(t) = x + \int_0^t \left[r(s)X(s) + Z^\top \sigma^{-1}(s)\{\mu(s) - r(s)\mathbf{1}\}\right] \mathrm{d}s + \int_0^t Z^\top \mathrm{d}W(s).$$

By Itô's formula we have that

$$|X(t)|^2 = |X(T)|^2 - 2\int_t^T X(s)\mathrm{d}X(s) - \int_t^T |Z(s)|^2 \mathrm{d}s.$$

Using Lemma 3.3.1 we see that $\int_0^\cdot X\pi^\top \sigma \mathrm{d}W$ is a martingale, and so taking the expectation in the above expression and omitting the time arguments, we get:

$$\begin{aligned} \mathbb{E}\left(|X(t)|^2 + \int_t^T |Z|^2 \mathrm{d}s\right) &= \mathbb{E}\left(|X(T)|^2 - 2\int_t^T r|X|^2 \mathrm{d}s - 2\int_t^T XZ^\top \sigma^{-1}\{\mu - r\mathbf{1}\}\mathrm{d}s\right) \\ &\leq \mathbb{E}\left(|X(T)|^2 + 2\|r\|_\infty \int_t^T |X|^2 \mathrm{d}s + 2\|\sigma^{-1}\{\mu - r\mathbf{1}\}\|_\infty \int_t^T |X||Z|\mathrm{d}s\right) \\ &\leq \mathbb{E}\left(|X(T)|^2 + 2\left(\|r\|_\infty + \|\sigma^{-1}\{\mu - r\mathbf{1}\}\|_\infty^2\right)\int_t^T |X|^2 \mathrm{d}s + \int_t^T \frac{|Z|^2}{2}\mathrm{d}s\right), \end{aligned}$$

from which

$$\mathbb{E}\left(|X(t)|^2 + \tfrac{1}{2}\int_t^T |Z|^2 \mathrm{d}s\right) \leq \mathbb{E}\left(|X(T)|^2 + 2\left(\|r\|_\infty + \|\sigma^{-1}\{\mu - r\mathbf{1}\}\|_\infty^2\right)\int_t^T |X|^2 \mathrm{d}s\right). \tag{3.5.19}$$

Since the above inequality implies that

$$\mathbb{E}\left(|X(t)|^2\right) \leq \mathbb{E}\left(|X(T)|^2 + 2\left(\|r\|_\infty + \|\sigma^{-1}\{\mu - r\mathbf{1}\}\|_\infty^2\right)\int_t^T |X|^2 \mathrm{d}s\right),$$

by Gronwall's Lemma we obtain that

$$\mathbb{E}\left(|X(t)|^2\right) \leq \mathbb{E}\left(|X(T)|^2\right)e^{2\left(\|r\|_\infty + \|\sigma^{-1}\{\mu - r\mathbf{1}\}\|_\infty^2\right)T}, \tag{3.5.20}$$

and the first estimate in (3.5.18) follows from estimate (3.5.19), Fubini's Theorem, the definition of Z and the fact that $\sigma\sigma^\top \geq \delta I_{d\times d}$. Finally, the second estimate in (3.5.18) is a consequence of (3.5.20) and Fubini's Theorem. ■

For $P \in \mathcal{P}_4$ let us write the dynamic constraint (3.5.16) as $G(X, \pi, P) = 0$ with

$$G(X, \pi, P) = x + \int_0^\cdot \left[r(t)X(t) + \pi(t)^\top\{\mu(t) - r(t)\mathbf{1}\}\right] \mathrm{d}t + \int_0^\cdot \pi(t)^\top \sigma(t)\mathrm{d}W(t) - X(\cdot),$$

and further consider $\hat{G}(X,\pi,P) = (G(X,\pi,P), \mathbb{E}[X(T)] - A)$. Let us prove first:

Lemma 3.5.4 *The function $\hat{G} : \mathcal{I}^1 \times (L^{2,2}_{\mathbb{F}})^d \times \mathcal{P}_4 \mapsto \mathcal{I}^1 \times \mathbb{R}$ is continuously Fréchet differentiable. Furthermore, if $P \in \hat{\mathcal{P}}_4$, then $D_{(x,\pi)}\hat{G}(X,\pi,P) : \mathcal{I}^1 \times (L^{2,2}_{\mathbb{F}})^d \mapsto \mathcal{I}^1 \times \mathbb{R}$ is onto.*

Proof. The Fréchet differentiability of $\hat{G}$ can be proved following exactly the same lines of the proof in Lemma 3.5.1 and using that the second component of $\hat{G}$ is a continuous linear functional. For the surjectivity claim, suppose that $P \in \hat{\mathcal{P}}_4$ and that we are given $Y \in \mathcal{I}^1$ and $\xi \in \mathbb{R}$. Then we need to find $(Z,\nu) \in \mathcal{I}^1 \times (L^{2,2}_{\mathbb{F}})^d$ such that:

$$
\begin{aligned}
-Z(\cdot) + \int_0^{\cdot} \left[rZ + \nu^\top(\mu - r\mathbf{1})\right] \mathrm{d}t + \int_0^{\cdot} \nu^\top \sigma \mathrm{d}W(t) &= Y_0 + \int_0^{\cdot} Y_1 \mathrm{d}t + \int_0^{\cdot} Y_2 \mathrm{d}W(t), \\
\mathbb{E}[Z(T)] &= \xi.
\end{aligned}
\tag{3.5.21}
$$

Let $i \in \{1,\dots,d\}$ be such that $\kappa := \mathbb{E}\left(\int_0^T [\mu_i(t) - r(t)]\right) \mathrm{d}t \neq 0$. Then, consider the portfolio ν with $\nu^j = 0$ for $j \neq i$ and

$$
\nu^i(t) := \left(\frac{\xi + e^{\int_0^T r(t)\mathrm{d}t}\left[Y_0 + \mathbb{E}\left(\int_0^T e^{-\int_0^t r(s)\mathrm{d}s} Y_1(t)\mathrm{d}t\right)\right]}{e^{\int_0^T r(t)\mathrm{d}t}\kappa} \right) e^{\int_0^t r(s)\mathrm{d}s}.
$$

Then, defining $Z \in \mathcal{I}^1$ as the solution of

$$
\begin{aligned}
dZ(t) &= \left[r(t)Z(t) + \nu^\top(\mu - r\mathbf{1}) - Y^1(t)\right] \mathrm{d}t + \left[\nu^\top \sigma - Y^2(t)\right] \mathrm{d}W(t), \\
Z(0) &= -Y_0,
\end{aligned}
$$

we easily check that (Z,ν) satisfies (3.5.21). ■

We now show that problem (MVP) is attained. From here onwards $P := (x, r, A, \mu, \sigma)$ in $\hat{\mathcal{P}}_4$ will denote a tuple of (reference, nominal) parameters. We denote by $v(P)$ the value of (MVP) under parameters P.

Lemma 3.5.5 *We have that $v(P) < \infty$, and further this value is attained at a unique feasible pair $(X[P], \pi[P])$. Moreover, there exists a unique weak-Pontryagin multiplier*

$$
(p[P], q[P], \lambda_E[P]) \in \mathcal{I} \times (L^{2,2}_{\mathbb{F}})^{1\times d} \times \mathbb{R},
$$

satisfying:

$$
\begin{aligned}
\mathrm{d}p[P](t) &= -r(t)p[P](t)\mathrm{d}t + q[P](t)\mathrm{d}W(t) \quad \text{for all } t \in]0,T[, \\
p[P](T) &= 2[X[P](T) - A] + \lambda_E[P] \quad \textit{a.s. in } \Omega, \\
p[P](t,\omega)(\mu(t,\omega) - r(t)\mathbf{1}) &= -\sigma(t,\omega)(q[P](t,\omega))^\top \quad \text{a.s. in } [0,T] \times \Omega.
\end{aligned}
\tag{3.5.22}
$$

Proof. For the finiteness of $v(P)$ it suffices to prove that the feasible set is non-empty. Indeed, by (3.5.17) there is an i such that $\mathbb{E}[\int_0^T(\mu^i(t)-r(t))\mathrm{d}t]\neq 0$. Therefore, as in the proof of Lemma 3.5.4, we may build the portfolio π having 0 in every coordinate except for the i-th one, which is set to

$$\left(\frac{A\exp\{-\int_0^T r(t)\mathrm{d}t\}-x}{\mathbb{E}[\int_0^T(\mu^i(t)-r(t))\mathrm{d}t]}\right)e^{\int_0^{\cdot} r(t)\mathrm{d}t}.$$

We easily see that the corresponding wealth process has expected return equal to A at time T and so it is feasible. Suppose now that (X^1,π^1) and (X^2,π^2) attain $v(P)$. This implies that $\mathbb{E}\left[(X^1(T))^2\right]=\mathbb{E}\left[(X^2(T))^2\right]$. If $X^1(T)$ were not almost surely equal to $X^2(T)$, by strict convexity of $Z\in L^2_{\mathcal{F}_T}\mapsto\mathbb{E}[Z^2]$ we would get that the pair $\frac{1}{2}(X^1+X^2,\pi^1+\pi^2)$ is feasible and induces a strictly smaller value of the objective function, yielding a contradiction. Calling now

$$\hat{X}(\cdot):=X^1(\cdot)-X^2(\cdot)=\int_0^{\cdot}\{r(X^1-X^2)+(\pi^1-\pi^2)^\top(\mu-r\mathbf{1})\}\mathrm{d}t+\int_0^{\cdot}(\pi^1-\pi^2)^\top\sigma\mathrm{d}W(t),$$

we see that $\hat{X}(T)=0$ and from Lemma 3.5.3 that $\pi^1-\pi^2\equiv 0$ and thus $\hat{X}(\cdot)\equiv 0$, and so that X^1 and X^2 are indistinguishable. For attainability, suppose first that (X^k,π^k) is a feasible optimizing sequence. We then know that $\mathbb{E}\left([X^k(T)]^2\right)$ is bounded. By Lemma 3.5.3 we get that π^k is bounded in $(L^{2,2}_{\mathbb{F}})^d$ and X^k is bounded in $L^{2,2}_{\mathbb{F}}$. Therefore, there exist $\pi\in(L^{2,2}_{\mathbb{F}})^d$, $\hat{X}\in L^{2,2}_{\mathbb{F}}$ such that, up to some subsequence, (X^k,π^k) converges weakly to $(\hat{X},\pi)$ in $L^{2,2}_{\mathbb{F}}\times(L^{2,2}_{\mathbb{F}})^d$. Moreover, since in $L^{2,2}_{\mathbb{F}}$ we have that X^k_1 converges weakly to $r\hat{X}+\pi^\top(\mu-r\mathbf{1})$ and X^k_2 converges weakly to $\pi^\top\sigma$, we obtain that X^k converges weakly in $\mathcal{I}^1$ to

$$X(\cdot):=x+\int_0^{\cdot}\left[r\hat{X}+\pi^\top(\mu-r\mathbf{1})\right]\mathrm{d}t+\int_0^{\cdot}\pi^\top\sigma\mathrm{d}W(t).$$

Therefore, using that $\mathcal{I}$ is injected continuously in $L^{2,2}_{\mathbb{F}}$ by Proposition 3.5.1(i), uniqueness of the weak limit implies that $\hat{X}=X$. Moreover, using Proposition 3.5.1(i) again we see that $\mathbb{E}[X^k(T)]=A$ passes to the limit and we obtain that (X,π) is a feasible pair. Since the cost function is convex and strongly continuous we have that it is l.s.c. with respect to the weak convergence in $\mathcal{I}^1$, which implies that (X,π) is the optimal pair. Finally, the existence and uniqueness of the weak-Pontryagin multiplier $(p[P],q[P],\lambda_E[P])$ is a direct consequence of Theorem 3.5.2, Remark 3.5.2(i) and Lemma 3.5.4. Using (3.4.11), it is straightforward to see that $(p[P],q[P],\lambda_E[P])$ satisfies (3.5.22). ■

In order to simplify the sensitivity analysis, we use a change of variables that reduces the number of parameters. We let $X'(\cdot):=e^{-\int_0^{\cdot} r\mathrm{d}t}X(\cdot)-Ae^{-\int_0^T r(t)\mathrm{d}t}$ and for the portfolio variables we define the new ones by $\pi'(\cdot)=e^{-\int_0^{\cdot} r\mathrm{d}s}\pi(\cdot)$. With this change of variables, we easily see that for $P'=(x-Ae^{-\int_0^T r(t)\mathrm{d}t},0,0,\mu-r\mathbf{1},\sigma)$ we have the identity

$$v(P) = e^{2\int_0^T r\mathrm{d}s} v(P'). \tag{3.5.23}$$

Moreover, $(\bar{X}, \bar{\pi}, \bar{p}, \bar{q}, \bar{\lambda}_E) = (X[P], \pi[P], p[P], q[P], \lambda_E[P])$ if and only if

$$\begin{aligned}
(X[P'], \pi[P']) &= \left(e^{-\int_0^{\cdot} r\mathrm{d}t} \bar{X}(\cdot) - Ae^{-\int_0^T r(t)\mathrm{d}t}, e^{-\int_0^{\cdot} r\mathrm{d}t} \bar{\pi} \right) \\
(p[P'], q[P'], \lambda_E[P']) &= \left(e^{\int_0^{\cdot} r\mathrm{d}t - 2\int_0^T r\mathrm{d}t} \bar{p},\ e^{\int_0^{\cdot} r\mathrm{d}t - 2\int_0^T r\mathrm{d}t} \bar{q},\ e^{-\int_0^T r\mathrm{d}t} \bar{\lambda}_E \right).
\end{aligned} \tag{3.5.24}$$

Therefore, in the following we will consider general perturbations with respect to the initial condition, the drift and diffusion coefficients, and for ease of notation we will write the value function only in terms of these parameters. That is, we shall assume that $r \equiv 0$, $A = 0$ and consider perturbed parameters of the form $P(k) := (x^k, \mu^k, \sigma^k)$. In the end of this section we shall undo the above change of variables and analyse the full original problem.

We will repeatedly use the notation

$$\begin{aligned}
(X^k, \pi^k, p^k, q^k, \lambda^k) &:= (X[P(k)], \pi[P(k)], p[P(k)], q[P(k)], \lambda_E[P(k)]), \\
(\bar{X}, \bar{\pi}, \bar{p}, \bar{q}, \bar{\lambda}_E) &:= (X[P], \pi[P], p[P], q[P], \lambda_E[P]),
\end{aligned}$$

We now prove a stability result, essential to our analysis.

Proposition 3.5.3 *For any sequence $P(k) \to P$ we have that $v(P(k)) \to v(P)$ and further*

$$\left(X^k, \pi^k, p^k, q^k, \lambda^k\right) \to \left(\bar{X}, \bar{\pi}, \bar{p}, \bar{q}, \bar{\lambda}_E\right),$$

strongly in $\mathcal{I}^1 \times (L^{2,2}_{\mathbb{F}})^d \times \mathcal{I}^1 \times (L^{2,2}_{\mathbb{F}})^{1,d} \times \mathbb{R}$.

Proof. First note that, since $P \in \hat{\mathcal{P}}_4$, there is a coordinate i (which we fix) such that $\left[\mathbb{E}(\int_0^T \mu^i(t)\mathrm{d}t)\right]^2 > 0$. This implies that, for k large enough,

$$\left[\mathbb{E}\left(\int_0^T (\mu^i)^k(t)\mathrm{d}t\right)\right]^2 \geq \frac{1}{2}\left[\mathbb{E}\left(\int_0^T \mu^i(t)\mathrm{d}t\right)\right]^2 > 0$$

and so the portfolios with i-th component equal to $-x^k/\mathbb{E}(\int_0^T (\mu^i)^k\mathrm{d}t)$ (and zero in the remaining ones) are feasible for $(MVP(k))$. Using these feasible portfolios, we easily get the existence of $K > 0$ (independent of k) such that $v(P(k)) = \mathbb{E}[X^k(T)^2] \leq K$ and thus by Lemma 3.5.3 we obtain that π^k is bounded in $(L^{2,2}_{\mathbb{F}})^d$.

Now, consider first only those k such that $v(P(k)) \geq v(P)$ and define a portfolio ν^k equals to $\bar{\pi}$ except for the i-th coordinate where it equals $\bar{\pi}^i + z^k$, with

$$z^k := \frac{-x^k - \mathbb{E}(\int_0^T \bar{\pi}^\top \mu^k\mathrm{d}t)}{\mathbb{E}(\int_0^T (\mu^i)^k\mathrm{d}t)}.$$

Calling $Z^k(\cdot) = x^k + \int_0^{\cdot}(\nu^k)^\top \mu^k \mathrm{d}t + \int_0^{\cdot}(\nu^k)^\top \sigma^k \mathrm{d}W(t)$, we easily check that (Z^k, ν^k) is feasible for $(MVP(k))$ and, since $z^k \to 0$, we have that $\mathbb{E}[(Z^k(T))^2] \to \mathbb{E}[(\bar{X}(T))^2]$. Hence, for any $\epsilon > 0$ and k large enough we obtain that $|v(P(k)) - v(P)| \leq v(P(k)) + \epsilon - \mathbb{E}[(Z^k(T))^2] \leq \epsilon$. On the other hand, by considering those k such that $v(P(k)) \leq v(P)$, with a similar manner we can construct out of π^k a new portfolio ξ^k obtained by modification of π^k's i-th component in a way that it becomes feasible for the unperturbed problem. More precisely, it suffices to set $(\xi^j)^k = (\pi^j)^k$ for $j \neq i$ and $(\xi^i)^k := (\pi^i)^k + \hat{z}^k$, where

$$\hat{z}^k := \frac{-x - \mathbb{E}(\int_0^T (\pi^k)^\top \mu \mathrm{d}t)}{\mathbb{E}(\int_0^T \mu^i \mathrm{d}t)} = \frac{x^k - x - \mathbb{E}(\int_0^T (\pi^k)^\top [\mu - \mu^k]\mathrm{d}t)}{\mathbb{E}(\int_0^T \mu^i \mathrm{d}t)}.$$

Since π^k is bounded in $(L^{2,2}_{\mathbb{F}})^d$ we obtain that $\hat{z}^k \to 0$ and, as before, we get that for every $\epsilon > 0$ and n large enough, $|v(P(k)) - v(P)| \leq \epsilon$, which proves convergence of the value functions.

Now let π be any weak limit point of π^k in $(L^{2,2}_{\mathbb{F}})^d$. Since, for $(y, Y, Z) \in \mathbb{R} \times L^{2,2}_{\mathbb{F}} \times (L^{2,2}_{\mathbb{F}})^d$

$$\langle X^k, (y, Y, Z)\rangle_{\mathcal{I}} = x^k y + \mathbb{E}\left(\int_0^T Y(\pi^k)^\top \mu^k \mathrm{d}t\right) + \mathbb{E}\left(\int_0^T (\pi^k)^\top \sigma^k Z \mathrm{d}t\right),$$

we get that, except for some subsequence, $\langle X^k, (y, Y, Z)\rangle_{\mathcal{I}} \to \langle X, (y, Y, Z)\rangle_{\mathcal{I}}$, where $X(\cdot) = x + \int_0^{\cdot} \pi^\top \mu \mathrm{d}t + \int_0^{\cdot} \pi^\top \sigma \mathrm{d}W(t)$, and thus $X^k \to X$ weakly in $\mathcal{I}^1$. Noticing that $\mathbb{E}(X(T)) = 0$, and by virtue of convergence of the value functions, we have similarly as in Lemma 3.5.5 that $(X, \pi) = (\bar{X}, \bar{\pi})$. By Proposition 3.5.1(i) we see that $X^k(T)$ converges weakly in $L^2_{\mathcal{F}_T}$ to $\bar{X}(T)$ and using that $\mathbb{E}([X^k(T)]^2) \to \mathbb{E}([\bar{X}(T)]^2)$ we obtain that $X^k(T) \to \bar{X}(T)$ strongly. Let us write $\hat{X}^k = x + \int_0^{\cdot}(\pi^k)^\top \mu \mathrm{d}t + \int_0^{\cdot}(\pi^k)^\top \sigma \mathrm{d}W(t)$. Then by Lemma 3.5.3:

$$\|\pi^k - \bar{\pi}\|_{2,2}^2 \leq C\mathbb{E}[(\hat{X}^k(T) - \bar{X}(T))^2],$$

where $C = C(\mu, \sigma) > 0$ is some positive constant. Now, we have that $\mathbb{E}[(\bar{X}(T) - X^k(T))^2]$ tends to zero, and

$$\mathbb{E}[(\hat{X}^k(T) - X^k(T))^2] \leq |x - x^k|^2 + T\|\pi^k\|_{2,2}^2 \left[\|\mu - \mu^k\|_{\infty,\infty}^2 + \|\sigma - \sigma^k\|_{\infty,\infty}^2\right],$$

which also tends to zero. We conclude with the triangle inequality that $\pi^k \to \bar{\pi}$ strongly in $(L^{2,2}_{\mathbb{F}})^d$. Finally, since

$$\|X^k - \bar{X}\|_{\mathcal{I}}^2 = |x - x^k|^2 + \|(\pi^k)^\top \mu^k - \bar{\pi}^\top \mu\|_{2,2}^2 + \|(\pi^k)^\top \sigma^k - \bar{\pi}^\top \sigma\|_{2,2}^2,$$

we conclude that $X^k \to \bar{X}$ strongly in $\mathcal{I}^1$. Now, for the weak-Pontryagin multipliers $(p^k, q^k, \lambda^k) \in \mathcal{I}^1 \times (L^{2,2}_{\mathcal{F}})^{1\times d} \times \mathbb{R}$, by (3.5.22) we have that:

$$\begin{aligned} \mathrm{d}p^k &= q^k \mathrm{d}W(t) \ \text{ for all } t \in]0,T[, \quad p^k(T) = 2X^k(T) + \lambda^k, \\ 0 &= p^k(t,\omega)\mu^k(t,\omega) + \sigma^k(t,\omega)(q^k(t,\omega))^\top, \quad \text{for a.a. } (t,\omega) \in [0,T] \times \Omega. \end{aligned}$$

We will show now that the λ^k are bounded uniformly in k. Define $P^k = p^k - \lambda^k$. Then we know that $P^k(T) = 2X^k(T)$ and $\mathrm{d}P^k(t) = q^k \mathrm{d}W(t)$. Since the $X^k(T)$ are $L^2_{\mathcal{F}_T}$-bounded, classical estimates for linear BSDEs imply that both q^k and P^k are bounded in $(L^{2,2}_{\mathbb{F}})^{1\times d}$ and $L^{2,2}_{\mathbb{F}}$, respectively. On the other hand we have that $(P^k + \lambda^k)\mu^k + \sigma^k(q^k)^\top = 0$, which proves that $\lambda^k \mu^k = -P^k\mu^k - \sigma^k(q^k)^\top$ and thus:

$$|\lambda^k| \|\mu^k\|_{2,2} = \|\lambda^k \mu^k\|_{2,2} \leq \|\mu^k\|_\infty \|P^k\|_{2,2} + \|\sigma^k\|_\infty \|q^k\|_{2,2}.$$

The right hand-side of the above expression is uniformly bounded by the nature of the perturbations we have, and the estimates we already had. Further, we check that $\|\mu^k\|_{2,2}$ is bounded away from zero since $\mu \not\equiv 0$ and thus λ^k is bounded. Take now any subsequence of (X^k, π^k, λ^k). Then, there exists $\hat{\lambda} \in \mathbb{R}$ such that, except for some subsequence, (X^k, π^k, λ^k) converges strongly to $(\bar{X}, \bar{\pi}, \hat{\lambda})$. This implies, by the classical estimates for linear BSDEs, that the corresponding (p^k, q^k) converge strongly in $L^{2,\infty}_{\mathbb{F}} \times (L^{2,2}_{\mathbb{F}})^{1\times d}$ to the solution (p, q) of

$$\mathrm{d}p(t) = q(t)\mathrm{d}W(t) \;\text{ for } t \in]0, T[, \quad p(T) = \bar{X}(T) + \hat{\lambda}.$$

Further, since $p^k(\cdot) = \lambda^k + \int_0^{\cdot} q^k \mathrm{d}W$, we have that $p^k \to p$ strongly in $\mathcal{I}^1$. Moreover, since $(q^k)^\top = -p^k(\sigma^k)^{-1}\mu^k$ converges in $(L^{2,2}_{\mathbb{F}})^{1,d}$ to $-p\sigma^{-1}\mu$ we conclude that $p\mu + \sigma q^\top = 0$. Therefore, by the uniqueness of the weak-Pontryagin multiplier in Lemma 3.5.5, we deduce that $(p, q, \hat{\lambda}) = (\bar{p}, \bar{q}, \bar{\lambda}_E)$. This proves that the whole sequence (p^k, q^k, λ^k) converges to $(\bar{p}, \bar{q}, \bar{\lambda}_E)$ strongly in $\mathcal{I}^1 \times (L^{2,2}_{\mathbb{F}})^{1,d} \times \mathbb{R}$. ■

By Lemma 3.5.4, Proposition 3.5.3 and arguing exactly as in the proof of Theorem 3.5.3, we have the following result:

Proposition 3.5.4 *The value function $v : \mathcal{P}_4 \mapsto \mathbb{R}$ is of class C^1 on $\hat{\mathcal{P}}_4$. Moreover, at every $P = (x, 0, 0, \mu, \sigma) \in \hat{\mathcal{P}}_4$ we have that*

$$D_{(x,\mu,\sigma)}v(P; \Delta P) = \bar{p}(0)\Delta x + \mathbb{E}\left[\int_0^T \bar{\pi}(t)^\top \Delta\mu(t)\bar{p}(t)\mathrm{d}t\right] + \mathbb{E}\left[\int_0^T \bar{\pi}(t)^\top \Delta\sigma(t)\bar{q}(t)^\top \mathrm{d}t\right], \tag{3.5.25}$$

where $\Delta P = (\Delta x, 0, 0, \Delta\mu, \Delta\sigma)$ and $(\bar{X}, \bar{\pi}, \bar{p}, \bar{q}, \bar{\lambda}_E) = (X[P], \pi[P], p[P], q[P], \lambda_E[P])$ is given by Lemma 3.5.5.

We now unwind the change of variables done in order to reduce the size of the parameter space. In this way we obtain sensitivities with respect to the initial capital, deterministic interest/saving rates, the desired return, the drift and the diffusion coefficients.

Theorem 3.5.4 *The value function $v : \mathcal{P}_4 \mapsto \mathbb{R}$ is C^1 on $\hat{\mathcal{P}}_4$. Moreover, at every*

$P = (x, r, A, \mu, \sigma) \in \hat{\mathcal{P}}_4$ we have that

$$
\begin{aligned}
&D_x v(P; \Delta x) = \bar{p}(0)\Delta x, \\
&D_r v(P; \Delta r) = \mathbb{E}\left(\int_0^T \bar{p}(t)(\bar{X}(t) - \bar{\pi}^\top \mathbf{1})\Delta r \mathrm{d}t \right), \\
&D_A v(P; \Delta A) = -\bar{\lambda}_E \Delta A, \\
&D_\mu v(P; \Delta \mu) = \mathbb{E}\left[\int_0^T \bar{\pi}(t)^\top \Delta\mu(t)\bar{p}(t)\mathrm{d}t \right], \\
&D_\sigma v(P; \Delta \sigma) = \mathbb{E}\left[\int_0^T \bar{\pi}(t)^\top \Delta\sigma(t)\bar{q}(t)^\top \mathrm{d}t \right],
\end{aligned}
\tag{3.5.26}
$$

where $(\bar{X}, \bar{\pi}, \bar{p}, \bar{q}, \bar{\lambda}_E) = (X[P], \pi[P], p[P], q[P], \lambda_E[P])$ is given by Lemma 3.5.5.

Proof. Since $(x, r, A, \mu, \sigma) \to (x - Ae^{-\int_0^T r(t)\mathrm{d}t}, 0, 0, \mu - r\mathbf{1}, \sigma)$ is C^1, we can apply the chain rule in (3.5.23). Therefore, by (3.5.24) and Proposition 3.5.4 we have that

$$
\begin{aligned}
D_x v(P) &= e^{2\int_0^T r(t)\mathrm{d}t}\, p[P'](0) = \bar{p}(0) \\
D_A v(P) &= e^{2\int_0^T r(t)\mathrm{d}t}\, (-e^{-\int_0^T r(t)\mathrm{d}t}) p[P'](0) = -e^{\int_0^T r(t)\mathrm{d}t}\lambda_E[P'] = -\bar{\lambda}_E \\
D_\mu v(P)\Delta\mu &= e^{2\int_0^T r(t)\mathrm{d}t}\, \mathbb{E}\left(\int_0^T e^{-\int_0^t r(s)\mathrm{d}s} \bar{\pi}(t)^\top \Delta\mu(t) e^{\int_0^t r(s)\mathrm{d}s - 2\int_0^T r(s)\mathrm{d}s} \bar{p}(t)\mathrm{d}t \right) \\
&= \mathbb{E}\left[\int_0^T \bar{\pi}(t)^\top \Delta\mu(t)\bar{p}(t)\mathrm{d}t \right] \\
D_\sigma v(P)\Delta\sigma &= e^{2\int_0^T r(t)\mathrm{d}t}\, \mathbb{E}\left(\int_0^T e^{-\int_0^t r(s)\mathrm{d}s} \bar{\pi}(t)^\top \Delta\sigma(t) e^{\int_0^t r(s)\mathrm{d}s - 2\int_0^T r(s)\mathrm{d}s} \bar{q}(t)^\top \mathrm{d}t \right) \\
&= \mathbb{E}\left[\int_0^T \bar{\pi}(t)^\top \Delta\sigma(t)\bar{q}(t)^\top \mathrm{d}t \right].
\end{aligned}
$$

Finally, setting $R(\cdot) := \int_0^\cdot \Delta r(t)\mathrm{d}t$ and using that $p[P'](0) = \lambda_E[P']$, we obtain

$$
\begin{aligned}
D_r v(P; \Delta r) &= 2R(T)v(P) + e^{\int_0^T r(t)\mathrm{d}t} p[P'](0)R(T)A \\
&\quad -e^{2\int_0^T r(t)\mathrm{d}t} \mathbb{E}\left[\int_0^T e^{-\int_0^t r(s)\mathrm{d}s} \bar{\pi}^\top \Delta r(t)\mathbf{1} e^{\int_0^t r(s)\mathrm{d}s - 2\int_0^T r(s)\mathrm{d}s} \bar{p}(t)\mathrm{d}t \right) \\
&= 2R(T)v(P) + \bar{\lambda}_E R(T)A - \mathbb{E}\left(\int_0^T \bar{\pi}^\top \Delta r(t)\mathbf{1}\bar{p}(t)\mathrm{d}t \right).
\end{aligned}
\tag{3.5.27}
$$

On the other hand,

$$
\mathbb{E}\left(\int_0^T \bar{X}(t)\bar{p}(t)\Delta r(t)\mathrm{d}t \right) = \mathbb{E}\left(R(T)\bar{X}(T)\bar{p}(T) - \int_0^T R(t)\mathrm{d}(\bar{X}(t)\bar{p}(t)) \right). \tag{3.5.28}
$$

By Itô's formula, we can write

$$
\begin{aligned}
\mathrm{d}(\bar{X}(t)\bar{p}(t)) &= \left[r\bar{X}(t)p(t) + \bar{\pi}^\top(\mu(t) - r(t)\mathbf{1})\bar{p}(t) - \bar{X}(t)r(t)\bar{p}(t) + \bar{\pi}(t)^\top\sigma(t)\bar{q}(t)^\top \right] \mathrm{d}t \\
&\quad + \left[\bar{X}(t)\bar{q} + \bar{p}(t)\bar{\pi}(t)^\top\sigma(t) \right] \mathrm{d}W(t).
\end{aligned}
$$

Since, by the third line in (3.5.22), $(\mu(t) - r(t)\mathbf{1})\bar{p}(t) = \sigma(t)\bar{q}(t)^\top$ we obtain with Lemma

3.3.1 that $\mathbb{E}\left(\int_0^T R(t)\mathrm{d}(\bar{X}(t)\bar{p}(t))\right) = 0$. Therefore, by (3.5.28) and the second line in (3.5.22), we get

$$\begin{aligned}\mathbb{E}\left(\int_0^T \bar{X}(t)\bar{p}(t)\Delta r(t)\mathrm{d}t\right) = \mathbb{E}\left(R(T)\bar{X}(T)\bar{p}(T)\right) &= R(T)\mathbb{E}\left(\bar{X}(T)[2(\bar{X}(T)-A)+\bar{\lambda}_E]\right), \\ &= 2R(T)v(P) + \bar{\lambda}_E R(T)A.\end{aligned} \tag{3.5.29}$$

The conclusion follows from (3.5.27) and (3.5.29). ■

Comparison with a known explicit result

We want to compare the theoretical sensitivities we obtained with those coming from a simplified model where an explicit solution is known. We choose to compare our results with the model in [Framstad et al., 2004, Example 4.1] (with null jump component). More precisely, we consider the (MVP) problem with $d=1$, $r\equiv 0$ and $\mu(\cdot):[0,T]\to\mathbb{R}$, $\sigma:[0,T]\to\mathbb{R}$ being deterministic bounded functions. Assuming that $\int_0^T \mu(t)\mathrm{d}t \neq 0$ and that σ is uniformly positive, problem (MVP) can be explicitly solved (see Framstad et al. [2004] for the details). Setting $\Sigma := \mu/\sigma$, the optimal portfolio, optimal states and adjoint states are given by

$$\begin{aligned}
\bar{X}(t) &= A + \frac{x-A}{e^{\int_0^T \Sigma^2(s)ds}-1}\left[e^{\int_0^T \Sigma^2(s)\mathrm{d}s}e^{-\int_0^t\left(\Sigma(s)\mathrm{d}W(s)+\frac{3}{2}\Sigma^2(s)\mathrm{d}s\right)}-1\right], \\
\bar{\pi}(t) &= \frac{(A-x)\mu(t)e^{\int_0^T \Sigma^2(s)\mathrm{d}s}}{\sigma(t)^2\left(e^{\int_0^T \Sigma^2(s)\mathrm{d}s}-1\right)}e^{-\int_0^t\left(\Sigma(s)\mathrm{d}W(s)+\frac{3}{2}\Sigma^2(s)\mathrm{d}s\right)}, && (3.5.30) \\
\bar{p}(t) &= \frac{2(x-A)}{e^{\int_0^T \Sigma^2(s)\mathrm{d}s}-1}e^{-\int_0^t\left(\Sigma(s)\mathrm{d}W(s)+\frac{1}{2}\Sigma^2(s)\mathrm{d}s\right)}, && (3.5.31) \\
\bar{q}(t) &= \frac{2(A-x)}{e^{\int_0^T \Sigma^2(s)\mathrm{d}s}-1}\Sigma(t)e^{-\int_0^t\left(\Sigma(s)\mathrm{d}W(s)+\frac{1}{2}\Sigma^2(s)\mathrm{d}s\right)}. && (3.5.32)
\end{aligned}$$

Thus, setting $P=(x,0,A,\mu,\sigma)$ and $\Delta P=(\Delta x,0,\Delta A,\Delta\mu,\Delta\sigma)$ from Theorem 3.5.4 we have

$$Dv(P;\Delta P) = \bar{p}(0)[\Delta x-\Delta A] + \mathbb{E}\left[\int_0^T \bar{p}(t)\bar{\pi}(t)\Delta\mu(t)\right] + \mathbb{E}\left[\int_0^T \bar{q}(t)\bar{\pi}(t)\Delta\sigma(t)\right].$$

If we assume that $\Delta\mu$ and $\Delta\sigma$ are deterministic, a brief computation then yields to:

$$D_x v(P;\Delta x) = \frac{2(x-A)\Delta x}{e^{\int_0^T \Sigma^2(s)\mathrm{d}s}-1}, \tag{3.5.33}$$

$$D_A v(P;\Delta A) = -\frac{2(x-A)\Delta A}{e^{\int_0^T \Sigma^2(s)\mathrm{d}s}-1}, \tag{3.5.34}$$

$$D_\mu v(P;\Delta\mu) = -\frac{2(A-x)^2 e^{\int_0^T \Sigma^2(s)\mathrm{d}s}}{\left(e^{\int_0^T \Sigma^2(s)\mathrm{d}s}-1\right)^2}\int_0^T \frac{\mu(t)\Delta\mu(t)}{\sigma(t)^2}\mathrm{d}t, \tag{3.5.35}$$

$$D_\sigma v(P;\Delta\sigma) = \frac{2(A-x)^2 e^{\int_0^T \Sigma^2(s)\mathrm{d}s}}{\left(e^{\int_0^T \Sigma^2(s)\mathrm{d}s}-1\right)^2}\int_0^T \frac{\mu(t)^2\Delta\sigma(t)}{\sigma(t)^3}\mathrm{d}t. \tag{3.5.36}$$

Since we know explicitly the solution, we can actually verify that

$$v(P) = \frac{(x-A)^2}{e^{\int_0^T \Sigma^2(s)\mathrm{d}s}-1},$$

and thus computing its derivatives we easily recover (3.5.33)-(3.5.36).

3.6 On a possible sensitivity analysis of the utility maximization problem

In this part we go back to the setting of expected utility maximization from terminal wealth in incomplete markets, as in Chapter 2. However, here we are interested in computing the several sensitivities with respect to the model parameters of a standard (non-robust) expected utility maximizing agent. This is in a sense a complementary approach to that of the mentioned chapter; whereas in the robust approach one seeks to hedge oneself against all possible (reasonable) inaccuracies in the price model, in the sensitivity approach one seeks to estimate how sensitive a particular price model is when one is led to changing its parameters slightly. We remark that the results obtained thus far in the present part of the thesis are not directly applicable to the utility maximization problem (for prices governed by a Brownian Motion) because of the presence of state constraints (requiring the wealth process to be bounded from below) and because even when working with square integrable portfolios (the control), an optimal one might just be locally square integrable and accordingly the optimal wealth may not belong to the Itô space as we defined it.

Let us take Proposition 2.5.6 as our starting point, and borrow all the notation used in section 2.5.1. The point of the mentioned proposition is that under the stated assumptions, the optimal utility from an initial wealth x and under the probability measure $\mathbb{P}$, can be computed as:

$$u_{\mathbb{P}}(x) = \sup_{Z\in L_J^+, J(Z)\le x} \mathbb{E}^{\mathbb{P}}(Z), \tag{3.6.1}$$

where $\{Z\in L_J^+, J(Z)\le x\}$ is a $\sigma(L_J, E_I)$, i.e. weak*, compact set.

The purpose of this section is to motivate how a sensitivity analysis can be performed by taking advantage of (3.6.1). For that matter let us consider a more specific, parametric, price model for this section, which generalizes the one we used for the mean-variance case. Suppose that the market consists of $d+1$ assets $S^0, S^1, \dots, S^d$ whose prices evolve under $\mathbb{P}$ as

$$\begin{aligned} \mathrm{d}S^0(t) &= rS^0(t), \quad \text{for } t \in [0,T], \quad S^0(0) = 1, \\ dS(t) &= \operatorname{diag}(S(t))\mu(t)\mathrm{d}t + \operatorname{diag}(S(t))\sigma(t)\mathrm{d}W(t) \quad \text{for} \quad t \in [0,T], \ S(0) = S_0 \in \mathbb{R}^d, \end{aligned} \tag{3.6.2}$$

where $S := (S^1, \dots, S^d)$ and W is a $\mathbb{P}$-brownian motion in $\mathbb{R}^n$. The precise properties on the processes $r \in L^{\infty,\infty}_{\mathbb{F}}$, $\mu \in (L^{\infty,\infty}_{\mathbb{F}})^d$ and $\sigma \in (L^{\infty,\infty}_{\mathbb{F}})^{d\times n}$ shall be given shortly and will imply that the financial market is arbitrage-free and complete (see e.g. [Karatzas and Shreve, 1998, Chapter 1,Theorem 4.2 and 6.6]).

Given an initial wealth $x \in \mathbb{R}$ and a *self-financing portfolio* $\pi \in (L^{2,2}_{\mathbb{F}})^d$ measured in units of wealth, the associated *wealth process* X is defined through the SDE:

$$\begin{aligned} \mathrm{d}X^\pi(t) &= \{r(t)X^\pi(t) + \pi(t)^\top(\mu(t) - r(t)\mathbf{1})\}\mathrm{d}t + \pi(t)^\top\sigma(t)\mathrm{d}W(t) \quad \text{for all} \quad t \in [0,T], \\ X^\pi(0) &= x. \end{aligned} \tag{3.6.3}$$

Here we face a modelling decision. If we want to study the model under parameters $(r^k, \mu^k, \sigma^k) := (r + \tau^k\Delta r, \mu + \tau^k\Delta\mu, \sigma + \tau^k\Delta\sigma)$, there are at least two options. One, which we call the *strong variant/formulation* is to consider a new process S^k like that of S but under the new parameters, so that the possible wealth precesses are of the form:

$$\begin{aligned} \mathrm{d}X^{\pi,k}(t) &= \{r^k(t)X^{\pi,k}(t) + \pi(t)^\top(\mu^k(t) - r^k(t)\mathbf{1})\}\mathrm{d}t + \pi(t)^\top\sigma^k(t)\mathrm{d}W(t) \quad , \\ X^{\pi,k}(0) &= x. \end{aligned} \tag{Prob(k)}$$

and the perturbed problem becomes (we use the s to denote *strong*):

$$u^s(r^k, \mu^k, \sigma^k) = \sup_{\substack{\pi \in L^2_{\mathbb{F}} \\ X^{\pi,k} \text{ as in } (Prob(k)) \\ X^{\pi,k}(\cdot) \geq 0}} \mathbb{E}^{\mathbb{P}}\left[U(X^{\pi,k}(T))\right]. \tag{3.6.4}$$

On the other hand, if we call $\lambda = \sigma^\top(\sigma\sigma^\top)^{-1}(\mu - r\mathbf{1})$ we get that the dynamics (3.6.3) can be expressed as

$$\begin{aligned} \mathrm{d}X^\pi(t) &= r(t)X^\pi(t)\mathrm{d}t + \pi(t)^\top\sigma\left[\lambda(t)\mathrm{d}t + \mathrm{d}W(t)\right] \quad \text{for all} \quad t \in [0,T], \\ X^\pi(0) &= x, \end{aligned} \tag{3.6.5}$$

therefore if we suppose that $(Prob(k))$ enjoys uniqueness in law (e.g. if $\pi(t)$ is a Lipschitz function of $X(t)$ and the model parameters are deterministic), then the above element X^π has under $\mathrm{d}\mathbb{Q}^k = \mathcal{E}\left(\int(\lambda^k - \lambda)(\cdot)^\top\mathrm{d}W(\cdot)\right)_T \mathrm{d}\mathbb{P}$ the same law as $X^{\pi,k}$ under $\mathbb{P}$, where $\lambda^k = (\sigma^k)^\top(\sigma^k(\sigma^k)^\top)^{-1}(\mu^k - r^k\mathbf{1})$ and $\mathcal{E}$ denotes the stochastic exponential. We would thus have that

$$\mathbb{E}^{\mathbb{P}}\left[U(X^{\pi,k}(T))\right] = \mathbb{E}^{\mathbb{Q}^k}\left[U(X^\pi(T))\right].$$

This leads us to consider the *weak variant*, where we move the parameters by changing the probability measure. Thus, we define the weak value-function as

$$u^w(r^k,\mu^k,\sigma^k) := \sup_{\substack{\pi\in L^2_{\mathbb{F}} \\ X^\pi \text{ as in } (3.6.3) \\ X^\pi(\cdot)\geq 0}} \mathbb{E}^{\mathbb{Q}^k}\left[U(X^\pi(T))\right].$$

In light of (3.6.1) we can further rewrite the above as:

$$u^w(r^k,\mu^k,\sigma^k) = \sup_{\substack{Z\in(L_J)^+ \\ J(Z)\leq x}} \mathbb{E}^{\mathbb{P}}\left[\exp\left\{\int_0^T (\lambda^k-\lambda)(t)^\top \mathrm{d}W(t) - \frac{1}{2}\int_0^T |(\lambda^k-\lambda)(t)|^2 \mathrm{d}t\right\} Z\right]. \tag{3.6.6}$$

Remark 3.6.1 *The value functions u^s, u^w defer in general. In Larsen and Žitković [2007a] the analysis of the continuity of u^s w.r.t. its arguments (actually λ) was performed. In Kardaras and Žitković [2011] the continuity of u^w (in a broader context) was analysed as well. Thanks to our rewriting of u^w in* (3.6.6)*, we will be able to deal with the analysis of the differentiability of this function w.r.t. all of its parameters. Let us remark that the differential (first and second order) analysis of the value function w.r.t. the initial wealth x was performed in e.g. Kramkov and Schachermayer [1999] and more specifically in Kramkov and Sîrbu [2006].*

We thus see that a sensitivity analysis of the weak value function is more tractable than for the strong one, since the perturbed parameters disappear from the state constraint and only enter through the stochastic exponential.

We will here perform a *sensitivity analysis in the weak formulation*, in the simplest possible setting, so as to motivate the type of results expected in a more general one. We make the following assumption, supposed to hold for the remaining of this section:

Assumption 16 *We have $n = d$ and suppose that the process σ is a.e. invertible with $|\sigma^{-1}(t,\omega)|$ uniformly bounded by a constant (market completeness), and take $U(x) = 2\sqrt{x}$ as our utility function.*

Let us rewrite $\lambda = \sigma^{-1}(\mu - r\mathbf{1})$. These parameters shall be kept fixed from now on. We also consider general, arbitrary parameters $(\bar r, \bar\mu, \bar\sigma)$ and call for simplicity:

$$u(\bar r,\bar\mu,\bar\sigma) := \sup_{\substack{\pi\in[L^2_{\mathbb{F}}]^d \\ X^\pi(0)=x,\ \mathrm{d}X^\pi \text{ as in } (3.6.3) \\ X^\pi(\cdot)\geq 0}} \mathbb{E}^{\bar{\mathbb{P}}}\left[U(X^\pi(T))\right], \tag{3.6.7}$$

where $\bar{\mathbb{P}}$ is the probability measure given by $\mathcal{E}(\int[\Lambda-\lambda]^\top \mathrm{d}W)\mathrm{d}\mathbb{P}$, and $\Lambda := \bar\sigma^{-1}(\bar\mu - \bar r\mathbf{1})$, so that if the parameters are close to the original ones then so is Λ to λ and also Novikov's condition implies that indeed $\bar{\mathbb{P}}$ is a probability measure.

Remark 3.6.2 *We easily see, in the jargon of section 2.5.1, that*

$$\begin{aligned} L_I &= \left\{ Z \in L^0(\mathcal{F}_T) : \mathbb{E}\left[\frac{Z^2}{\mathcal{E}(-\int \lambda \mathrm{d}W)_T} \right] < \infty \right\} \\ &= \left\{ Z \in L^0(\mathcal{F}_T) : \mathbb{E}\left[Z^2 \exp\left\{ \int_0^T \lambda^\top \mathrm{d}W + \frac{1}{2}\int_0^T |\lambda|^2 \mathrm{d}t \right\} \right] < \infty \right\}, \end{aligned}$$

and also $\|Z\|_I := \|Z\|_I^\ell = \left(\mathbb{E}\left[\frac{Z^2}{\mathcal{E}(-\int \lambda^\top \mathrm{d}W)_T} \right] \right)^{\frac{1}{2}}$. *Obviously* $L_I = E_I$ *in this context. Furthermore,*

$$L_J = \left\{ X \in L^0(\mathcal{F}_T) : \mathbb{E}\left[\mathcal{E}\left(- \int \lambda^\top \mathrm{d}W \right)_T X^2 \right] < \infty \right\},$$

and $\|X\|_J := \|X\|_J^a = 2\left(\mathbb{E}\left[\mathcal{E}(-\int \lambda^\top \mathrm{d}W)_T X^2\right]\right)^{\frac{1}{2}}$. *Thus* L_J *is just an* $L^2(\mathbb{Q})$ *space, where* $\mathrm{d}\mathbb{Q} := \mathcal{E}(-\int \lambda^\top \mathrm{d}W)_T \mathrm{d}\mathbb{P}$.

We go step by step. First let us define

$$g(\Lambda) = \mathcal{E}\left(\int [\Lambda - \lambda]^\top \mathrm{d}W \right)_T.$$

We shall prove continuity and Gâteaux differentiability of g, by repeatedly introducing SDEs and invoking Proposition 3.5.1. We hope that a related reasoning can be helpful in the future for different utility functions, at least of power type. For the different notions of directional differentiability see [Bonnans and Shapiro, 2000, Chapter 2.2.1].

Lemma 3.6.1 *The function* $g : [L_{\mathbb{F}}^{\infty,\infty}]^d \to L_I$ *is well defined and continuous w.r.t. the strong topologies. Moreover,* g *is Gâteaux differentiable, and we have that*

$$Dg(\Lambda)\Delta\lambda = \mathcal{E}\left(\int [\Lambda - \lambda]^\top \mathrm{d}W \right)_T \left\{ \int_0^T \Delta\lambda^\top \mathrm{d}W - \int_0^T \langle \Lambda - \lambda, \Delta\lambda \rangle \mathrm{d}t \right\}.$$

Proof. Let us first verify that g is well defined. Indeed:

$$\begin{aligned} \mathcal{E}\left(\int [\Lambda - \lambda]^\top \mathrm{d}W \right)_T \in L_I \quad &\iff \quad \mathbb{E}\left[\exp\left\{ \int_0^T (2\Lambda - \lambda)^\top \mathrm{d}W - \frac{1}{2}\int_0^T \left(2|\Lambda - \lambda|^2 - |\lambda|^2 \right) \mathrm{d}t \right\} \right] < \infty \\ &\iff \quad \mathbb{E}\left[\exp\left\{ \int_0^T (2\Lambda - \lambda)^\top \mathrm{d}W - \frac{1}{2}\int_0^T |2\Lambda - \lambda|^2 \mathrm{d}t \right\} K \right] < \infty, \end{aligned}$$

where $K := \exp\{\int_0^T |\Lambda|^2 \mathrm{d}t\}$. By assumption K is bounded by a constant, whereas

$$\mathbb{E}\left[\exp\left\{ \int_0^T (2\Lambda - \lambda)^\top \mathrm{d}W - \frac{1}{2}\int_0^T |2\Lambda - \lambda|^2 \mathrm{d}t \right\} \right] \leq 1,$$

since every positive local martingale is a super-martingale. Thus $g(\Lambda) \in L_I$.

Let us prove that g is continuous. Given Λ and L in $[L^{\infty,\infty}_{\mathbb{F}}]^d$ we compute by developing squares:

$$\|g(\Lambda)-g(L)\|_I^2=\mathbb{E}[A(T)]+\mathbb{E}[B(T)]-2\mathbb{E}[C(T)],$$

where we can check that

$$\begin{aligned}
A(T) &= \frac{g(\Lambda)^2}{\mathcal{E}\left(\int -\lambda^\top \mathrm{d}W\right)_T}\\
&= \exp\left\{\int_0^T (2\Lambda-\lambda)^\top \mathrm{d}W-\frac{1}{2}\int_0^T|2\Lambda-\lambda|^2\mathrm{d}t\right\}\exp\left\{\int_0^T|\Lambda|^2\mathrm{d}t\right\},\\
B(T) &= \frac{g(L)^2}{\mathcal{E}\left(\int -\lambda^\top \mathrm{d}W\right)_T}\\
&= \exp\left\{\int_0^T (2L-\lambda)^\top \mathrm{d}W-\frac{1}{2}\int_0^T|2L-\lambda|^2\mathrm{d}t\right\}\exp\left\{\int_0^T|L|^2\mathrm{d}t\right\},\\
C(T) &= \frac{g(\Lambda)g(L)}{\mathcal{E}\left(\int -\lambda^\top \mathrm{d}W\right)_T}\\
&= \exp\left\{\int_0^T (\Lambda+L-\lambda)^\top \mathrm{d}W-\frac{1}{2}\int_0^T[|L|^2+|\Lambda|^2+|\lambda|^2-2\langle\Lambda+L,\lambda\rangle]\mathrm{d}t\right\}.
\end{aligned}$$

By letting T be variable in the definitions of A,B,C, we get that the associated processes fulfil the following SDEs:

$$\begin{aligned}
\mathrm{d}A &= A\left\{|\Lambda|^2\mathrm{d}t+(2\Lambda-\lambda)^\top\mathrm{d}W\right\}\\
\mathrm{d}B &= B\left\{|L|^2\mathrm{d}t+(2L-\lambda)^\top\mathrm{d}W\right\}\\
\mathrm{d}C &= C\left\{\langle L,\Lambda\rangle\mathrm{d}t+(\Lambda+L-\lambda)^\top\mathrm{d}W\right\}.
\end{aligned}\tag{3.6.8}$$

By Proposition 3.5.1.ii we have that both B and C converge weakly to A as Itô processes (and in $L^{2,2}_{\mathbb{F}}$) as $L\to\Lambda$ strongly, and by Proposition 3.5.1.i we get that also their values at the final time T converge weakly (in $L^2_{\mathcal{F}_T}$), so in particular their expectations converge, proving that $\mathbb{E}[B(T)],\mathbb{E}[C(T)]\to\mathbb{E}[A(T)]$. This proves the desired continuity.

For the directional differentiability, we shall first prove that

$$\frac{g(\Lambda+\epsilon L)-g(\Lambda)}{\epsilon\sqrt{\mathcal{E}\left(-\int\lambda^\top\mathrm{d}W\right)_T}}\to\frac{\mathcal{E}\left(\int[\Lambda-\lambda]^\top\mathrm{d}W\right)_T\left\{\int_0^T L^\top\mathrm{d}W-\int_0^T\langle\Lambda-\lambda,L\rangle\mathrm{d}t\right\}}{\sqrt{\mathcal{E}\left(-\int\lambda^\top\mathrm{d}W\right)_T}},\tag{3.6.9}$$

weakly in $L^2_{\mathcal{F}_T}$, as $\epsilon\searrow 0$. Then we will prove that the L^2 norm of the l.h.s. converges to that of the r.h.s., and so as a consequence we get that the l.h.s. converges in $L^2_{\mathcal{F}_T}$ to the r.h.s., and thus $\frac{g(\Lambda+\epsilon L)-g(\Lambda)}{\epsilon}\to\mathcal{E}\left(\int[\Lambda-\lambda]^\top\mathrm{d}W\right)_T\left\{\int_0^T L^\top\mathrm{d}W-\int_0^T\langle\Lambda-\lambda,L\rangle\mathrm{d}t\right\}$

in L_I, proving the directional differentiability of g. The linearity of $Dg(\Lambda)(\cdot)$ and its continuity, which is consequence of a reasoning as in the proof of g being well-defined, then yield the full Gâteaux differentiability of g.

Let us prove the weak convergence in (3.6.9). Some simple computations yield that the process defined by the r.h.s. (taking a varying t instead of T, as before) is the solution to de SDE:

$$\mathrm{d}h = \{h[\Lambda - \lambda/2] + \exp(c)L\}^\top \mathrm{d}W + \{h[\langle\lambda,\Lambda\rangle/2 - |\lambda|/8] + \exp(c)\langle\lambda, L\rangle/2\}\mathrm{d}t,$$

where $c := \int_0^\cdot(\Lambda - \lambda/2)^\top \mathrm{d}W - \frac{1}{2}\int_0^\cdot[|\Lambda|^2 + |\lambda|^2/2 - 2\langle\Lambda,\lambda\rangle]\mathrm{d}t$. On the other hand, the process associated to the l.h.s. in (3.6.9) solves the following SDE:

$$\mathrm{d}h^\epsilon = \{h^\epsilon[\Lambda+\epsilon L-\lambda/2]+\exp(c)L\}^\top \mathrm{d}W+\{h^\epsilon[\langle\lambda,\Lambda+\epsilon L\rangle/2-|\lambda|/8]+\exp(c)\langle\lambda,L\rangle/2\}\mathrm{d}t.$$

By Proposition 3.5.1.ii we see that the process h^ϵ converges in the Itô sense to h and so in particular by part i of that result, we get the desired weak convergence in (3.6.9).

Now we prove that the $L^2_{\mathcal{F}_T}$ norm of the l.h.s. in (3.6.9) converges to that of the r.h.s. For that matter let us consider the notation of (3.6.8), where we rename B and C by B^ϵ and C^ϵ respectively and where we take the L variable there to be equal to the current $\Lambda + \epsilon L$. From this, we see that

$$\left[\frac{g(\Lambda+\epsilon L) - g(\Lambda)}{\epsilon\sqrt{\mathcal{E}\left(-\int\lambda^\top \mathrm{d}W\right)_T}}\right]^2 = \frac{A(T) + B^\epsilon(T) - 2C^\epsilon(T)}{\epsilon^2} =: R^\epsilon(T),$$

and we verify that viewed as a process,

$$\mathrm{d}R^\epsilon = \left\{R^\epsilon|\Lambda|^2 + B^\epsilon|L|^2 \ + 2\langle L,\Lambda\rangle\left(\frac{B^\epsilon - C^\epsilon}{\epsilon}\right)\right\}\mathrm{d}t + \left\{R^\epsilon[2\Lambda-\lambda] + 2L\left(\frac{B^\epsilon - C^\epsilon}{\epsilon}\right)\right\}^\top \mathrm{d}W.$$

On the other hand, we may verify that for

$$\left[\frac{\mathcal{E}\left(\int[\Lambda-\lambda]^\top \mathrm{d}W\right)_t\left\{\int_0^t L^\top \mathrm{d}W - \int_0^t\langle\Lambda-\lambda, L\rangle\mathrm{d}t\right\}}{\sqrt{\mathcal{E}\left(-\int\lambda^\top \mathrm{d}W\right)_t}}\right]^2 =: R(t),$$

the following SDE holds:

$$\mathrm{d}R = \left\{R|\Lambda|^2 + A|L|^2 \ + 2\langle L,\Lambda\rangle\Psi\right\}\mathrm{d}t + \{R[2\Lambda-\lambda] + 2L\Psi\}^\top \mathrm{d}W,$$

whereby $\Psi(t) := A(t)[\int_0^t L^\top \mathrm{d}W - \int_0^t\langle\Lambda-\lambda,L\rangle\mathrm{d}t]$. The $L^2_{\mathcal{F}_T}$-norm convergence that we want to prove is equivalent to the convergence $\mathbb{E}[R^\epsilon(T)] \to \mathbb{E}[R(T)]$. If we could prove that $R^\epsilon \to R$ weakly as Itô processes this would be a consequence of Proposition 3.5.1.i. On the other hand, by Proposition 3.5.1.ii and because we already know that $B^\epsilon \to A$ in $L^{2,2}_{\mathbb{F}}$, all we need to prove is that $\left(\frac{B^\epsilon - C^\epsilon}{\epsilon}\right) \to \Psi$ weakly in $L^{2,2}_{\mathbb{F}}$. Let us do this.

First we notice that Ψ satisfies:

$$\mathrm{d}\Psi = \{\Psi|\Lambda|^2 + A\langle\Lambda, L\rangle\}\mathrm{d}t + \{\Psi(2\Lambda - \lambda) + AL\}^\top \mathrm{d}W.$$

We next observe that calling $\Psi^\epsilon(t) := \left(\frac{B^\epsilon(t)-C^\epsilon(t)}{\epsilon}\right)$, then

$$\mathrm{d}\Psi^\epsilon = \{\Psi^\epsilon|\Lambda|^2 + (2B^\epsilon - C^\epsilon)\langle\Lambda, L\rangle\}\mathrm{d}t + \{\Psi^\epsilon(2\Lambda + \epsilon L - \lambda) + AB^\epsilon\}^\top \mathrm{d}W.$$

Because $B^\epsilon, C^\epsilon \to A$ in $L^{2,2}_{\mathbb{F}}$, we conclude that $\Psi^\epsilon \to \Psi$ in $L^{2,2}_{\mathbb{F}}$, by Proposition 3.5.1.ii and i. This completes the proof. ■

We can now prove by chain rule the following result. Some parts of the proof have nothing to do with our choice of utility function, pointing out that we may in the future extend the approach:

Proposition 3.6.1 *The function $f : [L^{\infty,\infty}_{\mathbb{F}}]^d \to \mathbb{R}^+$ defined by*

$$f(\Lambda) := \sup_{\substack{\pi\in[L^{2,2}_{\mathbb{F}}]^d \\ X^\pi(0)=x,\ \mathrm{d}X^\pi \text{ as in } (3.6.3) \\ X^\pi(\cdot)\geq 0}} \mathbb{E}\left[\mathcal{E}\left(\int[\Lambda-\lambda]^\top \mathrm{d}W\right)_T U(X^\pi(T))\right], \tag{3.6.10}$$

is continuous and Hadamard differentiable, whereby

$$Df(\Lambda, \Delta\Lambda) = \mathbb{E}\left[Z(\Lambda)\mathcal{E}\left(\int[\Lambda-\lambda]^\top \mathrm{d}W\right)_T \left\{\int_0^T \Delta\Lambda^\top \mathrm{d}W - \int_0^T \langle\Lambda-\lambda, \Delta\Lambda\rangle \mathrm{d}t\right\}\right],$$

and $Z(\Lambda)$ denotes the unique element attaining

$$\sup\left\{\mathbb{E}\left[\mathcal{E}\left(\int[\Lambda-\lambda]^\top \mathrm{d}W\right)_T Z\right] : Z \in \mathcal{L}^+_J, J(Z) \leq x\right\},$$

or equivalently, $Z(\Lambda) = U(X(\Lambda)_T)$ and $X(\Lambda)$ is the unique solution to (3.6.10).

Proof. We first write $f(\Lambda) = \sup\left\{\mathbb{E}\left[g(\Lambda)Z\right] : Z \in \mathcal{L}^+_J, J(Z) \leq x\right\}$, and denote $F : L_I \to \mathbb{R}$ the function $Y \mapsto \sup\left\{\mathbb{E}\left[YZ\right] : Z \in \mathcal{L}^+_J, J(Z) \leq x\right\}$, so that $f = F \circ g$. Let us observe that it would suffice to understand F acting on $E^{++}_I := \{Y \in E_I : Y > 0 \text{ a.s.}\}$ intersected with the sphere in L^1 only. We now verify that F is Lipschitz continuous. Indeed, by Proposition 2.5.4 we know that $\{Z \in \mathcal{L}^+_J : J(Z) \leq x\}$ is a weak* compact set in L_J, thus if we take Y_1, Y_2 and assume w.l.o.g. that $F(Y_1) \geq F(Y_2)$, we see:

$$|F(Y_1) - F(Y_2)| = F(Y_1) - F(Y_2) \leq \sup_{\{Z\in\mathcal{L}^+_J : J(Z)\leq x\}} \mathbb{E}(Z\{Y_1 - Y_2\}) \leq K\|Y_1 - Y_2\|_I,$$

where we used Proposition 2.5.3 and defined $K = \sup_{\{Z\in\mathcal{L}^+_J : J(Z)\leq x\}} \|Z\|_J < \infty$. By Lemma 3.6.1 we thus have that f is continuous.

Let us prove that F is a directionally differentiable function and actually Hadamard differentiable when restricted to E^{++}_I, and that $DF(Y, \Delta Y) = \mathbb{E}[Z(Y)\Delta Y]$, where $Z(Y)$ is the only element of the set $\mathcal{S}(Y) := \{Z \in \mathcal{L}^+_J : J(Z) \leq x, F(Y) = \mathbb{E}[ZY]\}$. Indeed, by Lemma 2.5.1.iii we know that $Z \in \mathcal{S}(Y) \Rightarrow \exists X \in \mathcal{X}(x)$ s.t. $Z \leq U(X(T))$, and so if $Y \in L_I^{++}$ the only possible elements in $\mathcal{S}(Y)$ are of the form $U(X(T))$, for some $X \in \mathcal{X}(x)$. However there exists exactly one element solving the problem

$$\sup_{X \in \mathcal{X}(x)} \mathbb{E}[YU(X(T))],$$

and so $\mathcal{S}(Y)$ must be a singleton. Indeed, the equivalent probability measure $\mathrm{d}\mathbb{P}^Y \propto Y\mathrm{d}\mathbb{P}$ induces a complete market whose unique martingale measure $\mathbb{Q}$ (the same as in the market under $\mathbb{P}$) has density equal to $Y^{-1}\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}$, whereby $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = \mathcal{E}\left(-\int \lambda \mathrm{d}W\right)_T$, and we verify that

$$\mathbb{E}^{\mathbb{P}^Y}[V(\beta Y^{-1}\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P})] = I(\beta Y) < \infty,$$

for every $\beta > 0$, since $Y \in E_I$. Therefore Assumption 9 holds and by Proposition 2.5.6 we have that $\mathcal{S}(Y)$ is a singleton.

From all this we can invoke [Bonnans and Shapiro, 2000, Proposition 2.121] and obtain that the subdifferential of F at any $Y \in E_I^{++}$ is a singleton and hence by [Bonnans and Shapiro, 2000, Proposition 2.126] F is Hadamard differentiable and $DF(Y)(\cdot) = \mathbb{E}[\cdot Z(Y)]$.

By the chain rule in [Bonnans and Shapiro, 2000, Proposition 2.47] and Lemma 3.6.1 we conclude that f is directionally differentiable and that its directional derivative is given by

$$Df(\Lambda, \Delta\Lambda) = \mathbb{E}\left[Z(\Lambda)\mathcal{E}\left(\int [\Lambda - \lambda]^\top \mathrm{d}W\right)_T \left\{\int_0^T \Delta\Lambda^\top \mathrm{d}W - \int_0^T \langle \Lambda - \lambda, \Delta\Lambda\rangle \mathrm{d}t\right\}\right],$$

with $Z(\Lambda)$ as in the statement of the present proposition. Using Proposition 2.5.3 we bound

$$|Df(\Lambda, \Delta\Lambda)| \leq \|Z(\Lambda)\|_J \left\|\mathcal{E}\left(\int [\Lambda - \lambda]^\top \mathrm{d}W\right)_T \left\{\int_0^T \Delta\Lambda^\top \mathrm{d}W - \int_0^T \langle \Lambda - \lambda, \Delta\Lambda\rangle \mathrm{d}t\right\}\right\|_I,$$

and since the first term in the r.h.s. is bounded by $1 + x$ and the second one is bounded whenever $\Delta\Lambda$ is taken in a bounded set (as in the proof of the Gâteaux differentiability of g in Lemma 3.6.1), we see that the directional derivative of f is continuous.

To close this proof we now show that f is Hadamard directionally differentiable. We first observe that

$$f(\Gamma) - f(\Lambda) = \int_0^1 Df(\Lambda + t[\Gamma - \Lambda])(\Gamma - \Lambda)\mathrm{d}t \leq \|\Gamma - \Lambda\|_\infty \sup_{0 \leq t \leq 1} \|Df(\Lambda + t[\Gamma - \Lambda])\|, \tag{3.6.11}$$

where the last norm is the operator norm (from $[L_{\mathbb{F}}^{\infty,\infty}]^d$ to $\mathbb{R}$). From here we shall prove that f is locally Lipschitz, and so an obvious modification of [Bonnans and Shapiro, 2000, Proposition 2.49] shows that f is indeed Hadamard directionally differentiable. Let us call $\tilde{\Lambda}$ an element in the vicinity of Λ (possibly of the form $\Lambda + t[\Gamma - \Lambda]$), and compute:

$$\begin{aligned} \|Df\left(\tilde{\Lambda}\right)\| &:= \sup_{\|\eta\|_\infty=1} \left| \mathbb{E}\left[Z(\tilde{\Lambda})\mathcal{E}\left(\int [\tilde{\Lambda}-\lambda]^\top \mathrm{d}W\right)_T \left\{ \int_0^T \eta^\top \mathrm{d}W - \int_0^T \langle \tilde{\Lambda}-\lambda, \eta\rangle \mathrm{d}t \right\}\right]\right| \\ &\leq \left\| Z(\tilde{\Lambda})\mathcal{E}\left(\int [\tilde{\Lambda}-\lambda]^\top \mathrm{d}W\right)_T \right\|_I \sup_{\|\eta\|_\infty=1} \left\| \int_0^T \eta^\top \mathrm{d}W - \int_0^T \langle \tilde{\Lambda}-\lambda, \eta\rangle \mathrm{d}t \right\|_J . \end{aligned}$$

Since by Girsanov's Theorem, $W^{\mathbb{Q}} := W + \int_0^\cdot \lambda \mathrm{d}t$ is a $\mathbb{Q}$-brownian motion, we have:

$$\begin{aligned} \left\| \int_0^T \eta^\top \{\mathrm{d}W - (\tilde{\Lambda}-\lambda)\mathrm{d}t\} \right\|_J^2 &= 4\mathbb{E}\left[\mathcal{E}\left(-\int \lambda^\top \mathrm{d}W\right)_T \left\{ \int_0^T \eta^\top \{\mathrm{d}W - (\tilde{\Lambda}-\lambda)\mathrm{d}t\} \right\}^2 \right] \\ &= 4\mathbb{E}^{\mathbb{Q}}\left[\left\{ \int_0^T \eta^\top \{\mathrm{d}W^{\mathbb{Q}} - \tilde{\Lambda}\mathrm{d}t\} \right\}^2 \right] \\ &\leq k(\|\tilde{\Lambda}\|_\infty), \end{aligned}$$

where $k(\cdot)$ is a quadratic function and we assumed $\|\eta\|_\infty = 1$. On the other hand, if we call $\mathrm{d}\tilde{\mathbb{P}} = \mathcal{E}\left(\int [\tilde{\Lambda}-\lambda]^\top \mathrm{d}W\right)_T \mathrm{d}\mathbb{P}$, we know (e.g. by Kramkov and Schachermayer [1999]) that the optimal wealth $\tilde{X}_T$ associated to the optimization problem under this measure, whose optimal value we denote $u_{\tilde{\Lambda}}(x)$, is given by

$$\tilde{X}_T = [U']^{-1}\left(u'_{\tilde{\Lambda}}(x) \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\tilde{\mathbb{P}}} \right) = \left(u'_{\tilde{\Lambda}}(x) \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\tilde{\mathbb{P}}} \right)^{-2},$$

and so we know that

$$Z(\tilde{\Lambda}) = U\left(\tilde{X}_T\right) = \left[\frac{2}{u'_{\tilde{\Lambda}}(x)} \right] \times \frac{\mathcal{E}\left(\int \{\tilde{\Lambda}-\lambda\}^\top \mathrm{d}W\right)_T}{\mathcal{E}\left(-\int \lambda^\top \mathrm{d}W\right)_T},$$

from which we compute

$$\begin{aligned} \left\| Z(\tilde{\Lambda})\mathcal{E}\left(\int [\tilde{\Lambda}-\lambda]^\top \mathrm{d}W\right)_T \right\|_I^2 &= \left[\frac{4}{(u'_{\tilde{\Lambda}}(x))^2} \right] \mathbb{E}\left[\frac{\mathcal{E}\left(\int \{\tilde{\Lambda}-\lambda\}^\top \mathrm{d}W\right)_T^4}{\mathcal{E}\left(-\int \lambda^\top \mathrm{d}W\right)_T^3} \right] \\ &= \left[\frac{4}{(u'_{\tilde{\Lambda}}(x))^2} \right] \mathbb{E}\left[\mathcal{E}\left(\int \{4\tilde{\Lambda}-\lambda\}^\top \mathrm{d}W\right)_T \exp\left[\int_0^T 6|\tilde{\Lambda}|^2 \right]\right] \\ &\leq \frac{K(\|\tilde{\Lambda}\|_\infty)}{(u'_{\tilde{\Lambda}}(x))^2}, \end{aligned}$$

whereby $K(\cdot)$ is a continuous function. Up to now we have proved that

$$\|Df\left(\tilde{\Lambda}\right)\|^2 \leq \frac{k(\|\tilde{\Lambda}\|_\infty)K(\|\tilde{\Lambda}\|_\infty)}{(u'_{\tilde{\Lambda}}(x))^2}.$$

Because $u_{\tilde\Lambda}(\cdot)$ is concave and strictly increasing we have:

$$u'_{\tilde\Lambda}(x) \geq u_{\tilde\Lambda}(x+1) - u_{\tilde\Lambda}(x) > 0,$$

so recalling that we already proved that $u_{\tilde\Lambda}(x) = f(\tilde\Lambda)$ is continuous w.r.t. $\tilde\Lambda$, and this is also valid for $u_{\tilde\Lambda}(x+1)$, we get that $u'_{\tilde\Lambda}(x) > 0$ on some vicinity of the original Λ, proving that $\|Df\left(\tilde\Lambda\right)\| \leq C$ in that neighbourhood (for some constant C) and hence by (3.6.11):

$$|f(\Gamma) - f(\Lambda)| \leq C\|\Gamma - \Lambda\|_\infty,$$

for any Γ near enough to Λ. This entails the locally Lipschitz character of f and finishes the proof. ■

We finally prove the desired sensitivities of the (weak) value function given in (3.6.7) with respect to its parameters:

Theorem 3.6.1 *Under Assumption 16 and calling*

$$\mathcal{P}_5 := \left\{(\bar r, \bar\mu, \bar\sigma) \in L^{\infty,\infty}_{\mathbb{F}} \times (L^{\infty,\infty}_{\mathbb{F}})^d \times (L^{\infty,\infty}_{\mathbb{F}})^{d\times d} : \sigma \textit{ is a.e. invertible and } \operatorname*{ess\,sup}_{t,\omega} |\sigma^{-1}| < \infty\right\}$$

we have that after fixing $(r, \mu, \sigma) \in \mathcal{P}_5$, the function

$$(\bar r, \bar\mu, \bar\sigma) \in \mathcal{P}_5 \mapsto u(\bar r, \bar\mu, \bar\sigma),$$

given in (3.6.7) *is continuous and Gâteaux differentiable, with directional derivatives given by:*

$$\begin{aligned}
D_r u(\bar r, \bar\mu, \bar\sigma)\Delta r &= \mathbb{E}\left[U(\bar X(T))L\left\{-\int_0^T [\bar\sigma^{-1}\Delta r\mathbf{1}]^\top \mathrm{d}W + \int_0^T \langle \Lambda - \lambda, \bar\sigma^{-1}\Delta r\mathbf{1}\rangle \mathrm{d}t\right\}\right]\\
D_\mu u(\bar r, \bar\mu, \bar\sigma)\Delta \mu &= \mathbb{E}\left[U(\bar X(T))L\left\{\int_0^T [\bar\sigma^{-1}\Delta \mu]^\top \mathrm{d}W - \int_0^T \langle \Lambda - \lambda, \bar\sigma^{-1}\Delta \mu\rangle \mathrm{d}t\right\}\right]\\
D_\sigma u(\bar r, \bar\mu, \bar\sigma)\Delta \sigma &= \mathbb{E}\left[U(\bar X(T))L\left\{-\int_0^T [\bar\sigma^{-1}\Delta \sigma\Lambda]^\top \mathrm{d}W + \int_0^T \langle \Lambda - \lambda, \bar\sigma^{-1}\Delta \sigma\Lambda\rangle \mathrm{d}t\right\}\right],
\end{aligned}$$

where $\lambda = \sigma^{-1}(\mu - r\mathbf{1})$, $\Lambda = \bar\sigma^{-1}(\bar\mu - \bar r\mathbf{1})$, $L := \mathcal{E}\left(\int [\Lambda - \lambda]^\top \mathrm{d}W\right)_T$ and $\bar X(T)$ is the unique optimal terminal wealth attaining $u(\bar r, \bar\mu, \bar\sigma)$.

Proof. We start by remarking that $\mathcal{P}_5$ is an open set of $L^{\infty,\infty}_{\mathbb{F}} \times (L^{\infty,\infty}_{\mathbb{F}})^d \times (L^{\infty,\infty}_{\mathbb{F}})^{d\times d}$. Next we notice that the function H given by $(\bar r, \bar\mu, \bar\sigma) \in \mathcal{P}_5 \mapsto \bar\sigma^{-1}(\bar\mu - \bar r\mathbf{1})$ is clearly continuous and Gâteaux differentiable with

$$DH(\bar r, \bar\mu, \bar\sigma)(\Delta r, \Delta\mu, \Delta\sigma) = -\bar\sigma^{-1}\Delta r\mathbf{1} + \bar\sigma^{-1}\Delta\mu - \bar\sigma^{-1}\Delta\sigma\bar\sigma^{-1}(\bar\mu - \bar r\mathbf{1}).$$

Finally we note that $u(\bar r, \bar\mu, \bar\sigma) = f \circ H(\bar r, \bar\mu, \bar\sigma)$, for f as in (3.6.10). By Proposition

3.6.1 we see that u is indeed continuous and the composition of a Hadamard and a directionally differentiable function. By [Bonnans and Shapiro, 2000, Proposition 2.47] we conclude that u is directionally differentiable, and these derivatives are as stated. Finally, as we did in the end of Proposition 3.6.1, we may use Proposition 2.5.3 to prove the continuity of the directional derivatives proving the Gâteaux differentiability. ■

Remark 3.6.3 *The extension of this result to more general utility functions and incomplete markets is an ongoing work.*

Remark 3.6.4 *The most meaningful message of the previous theorem is when we analyze the sensitivity of the weak value function around the nominal parameters themselves (r, μ, σ) (which amount to taking $(\bar{r}, \bar{\mu}, \bar{\sigma}) = (r, \mu, \sigma)$). In this case we get that*

$$Du(r, \mu, \sigma)(\Delta r, \Delta\mu, \Delta\sigma) = \mathbb{E}\left[U(X^*(T)) \int_0^T [\sigma^{-1}(\Delta\mu - \Delta r\mathbf{1} - \Delta\sigma\sigma^{-1}\{\mu - r\mathbf{1}\})]^\top \mathrm{d}W\right],$$

or equivalently

$$Du(\lambda)(\Delta\Lambda) = \mathbb{E}\left[U(X^*(T)) \int_0^T \Delta\Lambda^\top \mathrm{d}W\right],$$

where X^* is the optimal wealth under parameters (r, μ, σ).

We conclude by commenting informally for the last time about the relationship between the sensitivities computed from the weak and strong variant of the problem (the latter will be subject of future research). Let us take $r \equiv 0$ to simplify and consider u^s as in (3.6.4), where we write everything in terms of $\lambda = \sigma^{-1}\{\mu - r\mathbf{1}\}$ as usual. Eventhough the sensitivity analysis of the strong formulation is hard due to the presence of the parameters in the state constraint, we can reasonably conjecture (if anything like the "envelope" or "Danskin Theorem" is to hold for it, as well as a directional chain rule) that

$$Du^s(\lambda)\Delta\Lambda = \mathbb{E}\left[U'(X^*(T)) \int_0^T \pi^* \cdot \Delta\Lambda \mathrm{d}t\right],$$

where X^* is as in the previous remark and π is a corresponding optimal portfolio. The question is, how does this expression compare to that in Remark 3.6.4? If we suppose that both π^* and λ are deterministic, we may use Malliavin Calculus (we refer to Nualart [2006] for a thorough treatment) and obtain:

$$\begin{aligned}
\mathbb{E}\left[U(X^*(T)) \int_0^T \Delta\Lambda^\top \mathrm{d}W\right] &= \mathbb{E}\left[\int_0^T D_t[U(X^*(T))] \cdot \Delta\Lambda \mathrm{d}t\right] \\
&= \mathbb{E}\left[U'(X^*(T)) \int_0^T D_t[X^*(T)] \cdot \Delta\Lambda \mathrm{d}t\right] \\
&= \mathbb{E}\left[U'(X^*(T)) \int_0^T \pi^* \cdot \Delta\Lambda \mathrm{d}t\right],
\end{aligned}$$

where the first equality comes from the integration by parts formula, the second by the chain rule and the third because for Malliavin derivatives of deterministic integrands we have $D_t\left[\int_0^T \pi^* \mathrm{d}W\right] = \pi^* \mathbb{1}_{[0,T]}(t)$. We therefore see that the (candidate) sensitivity in the strong formulation then should coincide with the one in the weak formulation, at least if the parameters are deterministic as well as the optimal portfolio. Actually this should not be surprising, given that in this case we know that equation ($Prob(k)$) has uniqueness in law and so the weak and strong formulations have the same value at λ. It is nevertheless interesting that the perturbation direction $\Delta\Lambda$ may be taken random and yet still both sensitivities apparently coincide. The previous heuristic analysis also shows that in more general situations both sensitivities should differ, as the Malliavin derivative of $X^*(T)$ shall involve more terms.

4 Conditional Analysis and a Principal-Agent problem

4.1 Introduction

In this last chapter we shall analyse a so-called Principal-Agent problem under moral hazard, the main motivation of it being the case of delegated portfolio management under hidden or non-contractible action. We mainly follow the modelling framework of Ou-Yang [2003], but in the discrete-time case, and will constantly compare our results with theirs. Thanks to our assumptions we will be able to reduce the original dynamic stochastic optimization problem to a series of random, static ones, which unfortunately lack a certain convexity. A simple argument however will then reduce these problems to a new series of convex (strictly speaking concave, as we will be always maximizing here) unconstrained ones. With the aid of *Conditional Analysis* we are then able to solve these problems in a variety of situations and hence construct an optimal contract.

In section 4.2 we introduce the modelling framework and fix notation. Then in section 4.3 we define the way we model the preferences of both the Principal and the Agent. In section 4.4 thereafter we introduce the linear contracts that the Principal has access to and derive first formally and then rigorously the principle according to which the dynamic problem reduces to a sequence of static, potentially non-convex ones. We close that section with the result showing that these problems may be replaced by unconstrained, convex ones. Later in section 4.5 we survey some of the results in Conditional Analysis that we will need, as well as derive a few new ones. Then in section 4.6 we provide the most general attainability results we have, followed by section 4.7 and 4.8 where we specialize our analysis progressively. We close the chapter with 4.9 where we discuss some possible extensions and conclude.

4.2 On the model

Let us introduce the model we will be occupied with. We start by fixing a discrete-time setting, where the time-grid is given by $\{t_i = i\Delta t : i = 0, \dots, T/\Delta t\}$ and T (the trading horizon) is a finite deterministic time. We denote by Ω the set of states-of-the-world.

We introduce an $N-$dimensional, strictly positive, *discounted price* process P defined on Ω whose filtration we denote by $(\mathcal{F}_t)_t$, representing the prices of N tradable goods (we can think of stocks). The notation ΔP_{t+1} will be a short-hand for $P_{t+1} - P_t$ and $\Delta \tilde{P}_{t+1} = diag(P_t)^{-1}\Delta P_{t+1}$, whereby $diag(\cdot)$ denotes the natural diagonal matrix associated to the vector in its argument. The same notation applies for other processes different than P. We shall write $P_{0:t}$ to denote the path of the price process from time 0 to t. We take the convention that vectors are regarded as column ones.

At each time $t < T$ the Agent (he) will choose an N-dimensional strategy/effort-level/action A_t only known to him and adapted to the price filtration, representing the amount of money invested in each asset in the portfolio during the time interval $[t, t+1)$, and costing him $c(t, A_t)$ units of wealth, where $c(t, \cdot) : \mathbb{R}^N \to \mathbb{R}$ are given cost functions which we shall always assume convex (hence continuous). With this he generates a wealth increment proportional to $\Delta\tilde{P}$ by investing in the market. More concretely, if we denote by W_t^A the wealth accumulated by a self-financing portfolio up to time t, then:

$$W_T^A = W_0 + \Delta W_1^A + \cdots + \Delta W_T^A = W_0 + \sum_{t<T} A_t' \Delta \tilde{P}_{t+1}. \tag{4.2.1}$$

As for the Principal (she), we assume that she observes progressively the price process as well as the wealth W^A generated by the Agent. We remark that in principle this does not imply that the Principal observes Agent's decisions.

Regarding how the actors in the market interact, we suppose that the Principal will offer the Agent a contract, consisting of a linear combination of a fixed payment (contingent on the whole history of the price process) and a reward depending linearly on the wealth generated by Agent's portfolio, which is payable at the closing date T. The explicit expression of contracts will be introduced in the next sections, as well as the form of the preferences of the Agent and the Principal.

We finally suppose that we have a probability measure $\mathbb{P}$ defined on $(\Omega, \mathcal{F}_T)$. Let us for once say that equalities and inequalities are to be understood in the $\mathbb{P}$-almost sure sense, so for instance $X = Y$ (respect. $X \neq Y$) means that the two random variables agree (respect. disagree) on a set of full measure, unless otherwise stated.

On the economical side, we suppose that the market described by P and our filtered probability space is *free of arbitrage*, in the sense that:

$$\mathbb{P}\left(\sum A_t' \Delta P_{t+1} \geq 0\right) = 1 \Rightarrow \sum A_t' \Delta P_{t+1} = 0 \text{ a.s.},$$

and we do not assume market completeness.

4.3 On the preferences

Regarding the Agent's and Principal's "valuations" (i.e. the way they quantitatively assess the world), a natural assumption is that at time t they update theirs after observing respectively $\{P_s\}_{s\leq t}$ and $\{P_s, W_s^A\}_{s\leq t}$, reflecting the fact that the Agent only observes P and reacts accordingly, whereas the Principal observes both P and the output W^A progressively. Let us denote by $\mathcal{F}^A$ the natural filtration of (P, W^A) and remark that for the Agent both filtrations introduced so far coincide whereas for the Principal they differ unless she knew Agent's actions. From this we see that at time t both actors have a history-dependent utility functional. Thus Principal's and Agent's preferences are encoded by a stream of operators, which we shall call *family of utility functionals*, of the form:

$$U_t^a : L^0(\mathcal{F}_T) \to \underline{L}^0(\mathcal{F}_t) \quad \text{and} \quad U_t^p : L^0(\mathcal{F}_T^A) \to \underline{L}^0(\mathcal{F}_t^A),$$

whereby $L^0(\mathcal{F}_T)$ denotes the set of real-valued $\mathcal{F}_T-$measurable functions, and $\underline{L}^0(\mathcal{F}_t)$ the real and possibly $\{-\infty\}-$valued $\mathcal{F}_t-$measurable functions, and likewise for $\mathcal{F}^A$. This is in stark contrast with the von Neumann-Morgenstern representation (expected utility) and certainly generalizes in principle conditional expectation-based representations. We shall use the notation U^a referring to the Agent's utilities and U^p to the Principal's.

We now introduce several general (desirable) properties of utility functionals which we will use in Assumption 17 in identifying both U^a and U^p. Let us denote by $U_t : L^0(\mathcal{F}_T) \mapsto \underline{L}^0(\mathcal{F}_t)$ a generic family of utility functionals. The same concepts can be defined w.r.t. $\mathcal{F}^A$. Then $U := \{U_t\}_t$ is said to be:

- *normalized if* $U_t(0) = 0$,
- *proper if* $-\infty \not\equiv U_t(\cdot) < +\infty$,
- *monotone if* $U_t(X) \geq U_t(Y)$ whenever $X, Y \in L^0(\mathcal{F}_T)$ and $X \geq Y$ a.s.,
- *$\mathcal{F}_t$-conditionally concave if* $U_t(\lambda X + (1-\lambda)Y) \geq \lambda U_t(X) + (1-\lambda)U_t(Y)$ whenever $\lambda \in L^0(\mathcal{F}_t) \cap [0,1]$ and $X, Y \in L^0(\mathcal{F}_T)$,
- *$\mathcal{F}_t$-translation invariant if* $U_t(X+Y) = U_t(X) + Y$ whenever $X \in L^0(\mathcal{F}_T)$ and $Y \in L^0(\mathcal{F}_t)$,
- *time consistent if* $U_{t+1}(X) \geq U_{t+1}(Y) \Rightarrow U_t(X) \geq U_t(Y)$,

and the former hold for every $t \leq T$. Of course for a specific t all but the last property above can be defined for an operator $V : L^0(\mathcal{F}_T) \mapsto \underline{L}^0(\mathcal{F}_t)$.

We shall denote $dom(U_t) := \{X \in L^0(\mathcal{F}_T) : U_t(X) \in L^0(\mathcal{F}_t)\}$. It is well-known (see Cheridito et al. [a] or Filipović et al. [2012]) in particular that from these axioms follows the so-called *tower property* stating that $U_t(X) = U_t(U_{t+1}(X))$ whenever $X \in dom(U_{t+1})$, as well as the *local property* saying that $\mathbb{1}_A U_t(X) = \mathbb{1}_A U_t(Y)$ whenever $X, Y \in L^0(\mathcal{F}_T), A \in \mathcal{F}_t$ and $\mathbb{1}_A X = \mathbb{1}_A Y$. All of the previous properties will be referred to as the *usual properties/assumptions/conditions*.

For a discussion on these properties the reader may check Cheridito et al. [a]. We present some basic examples of such functionals:

Example 4.3.1 *We start with $\tilde{U}_t : L^\infty(\mathcal{F}_{t+1}) \mapsto L^\infty(\mathcal{F}_t)$ as given normalized and $\mathcal{F}_t-$translation invariant functionals, for which the extensions*

$$X \mapsto \lim_{n\to\infty} \lim_{m\to\infty} \tilde{U}_t([X \wedge n] \vee m),$$

are well defined between $\underline{L}(\mathcal{F}_{t+1})$ *and* $\underline{L}(\mathcal{F}_t)$*, and we keep calling them* $\tilde{U}_t$*. It is not difficult to see that* $U_t(X) := \tilde{U}_t \circ \tilde{U}_{t+1} \circ \cdots \circ \tilde{U}_T(X)$ *forms a time consistent and translation invariant family.*

Example 4.3.2 *Both the risk neutral family* $X \mapsto \mathbb{E}(X|\mathcal{F}_t)$ *and the entropic family* $X \mapsto -\frac{1}{\gamma}\log\left(\mathbb{E}(exp(-\gamma X)|\mathcal{F}_t)\right)$ *satisfy the usual assumptions (using the construction in the previous example). What is more, the family* $X \mapsto \operatorname{ess\,sup}_{s\in\mathbb{R}}\{s-\mathbb{E}[H(s-X)|\mathcal{F}_t]\}$ *satisfies these properties as well, provided* H *is a convex and increasing function with the property* $\sup_s[s - H(s)] = 0$*. The latter family of functionals is called Optimized Certainty Equivalent (this is actually their dual representation, see Cheridito et al. [a]) and we shall often refer to them. In Cheridito et al. [a] and the references therein the extension property of the previous example is discussed and for instance it is proved that optimized certainty equivalents allow for it.*

Before presenting the standing assumptions throughout this chapter we introduce some notation and terminology which will be often used in the sequel; given two sigma-algebras $\mathcal{K} \subset \tilde{\mathcal{K}}$ we denote $L^1_{\mathcal{K}}(\tilde{\mathcal{K}}) := \left\{Z \in L^0(\tilde{\mathcal{K}}) : \mathbb{E}[|Z||\mathcal{K}] \in L^0(\mathcal{K})\right\}$, and remark that clearly $L^1_{\mathcal{K}}(\tilde{\mathcal{K}}) = L^0(\mathcal{K})L^1(\tilde{\mathcal{K}})$ as sets. Also, we will say that a utility functional U is L^0-upper semicontinuous whenever at each $X \in dom(U)$ and for each sequence $X_n \to X$ almost surely, it holds that $\limsup U(X_n) \leq U(X)$ a.s. In section 4.5 a more topological notion of semicontinuity, related to $L^1_{\mathcal{K}}(\tilde{\mathcal{K}})$ (and more general spaces) will be given, and we shall call this notion $L^1_{\mathcal{K}}(\tilde{\mathcal{K}})-$semicontinuity.

We proceed to our standing Assumption in this work:

Assumption 17 *Both* U^a *and* U^p *satisfy the normalization, properness, monotonicity, conditional concavity, translation invariance and time consistency axioms (i.e. they satisfy the usual assumptions) with respect to* $\mathcal{F}$ *and* $\mathcal{F}^A$ *respectively. Furthermore, we assume that* $U^p_t(L^0(\mathcal{F}_T)) \subset \underline{L}^0(\mathcal{F}_t)$ *for each* t*.*

Remark 4.3.1 *The last point in the previous assumption means that* U^p *is* $\mathcal{F}$*-adapted over* $L^0(\mathcal{F}_T)$ *and implies that such restriction satisfies the usual assumptions w.r.t.* $\mathcal{F}$*.*

In the following we will always suppose Assumption 17 to hold. We move on now to the design of contracts.

4.4 On the contracts

The simplest contracts the Principal may offer the Agent are those (performance-dependent, linear) ones contingent on the possible realizations of W^A_T. Concretely, such a contract (or more exactly, a menu of payments) is defined as:

$$\bar{S} = \left\{A \mapsto \bar{S}(A) := \epsilon(P_{0:T}) + \beta W^A_T\right\},$$

where ϵ is a Borel-measurable functions and β is a constant. Thus the contract consists of a fixed (lump-side) payment ϵ which we may interpret as a financial derivative contingent only on the path of the price process (i.e. $\mathcal{F}_T$-measurable), plus a constant β times the closing value of the portfolio W_T^A. Thus we may call β the sensitivity-to-wealth component of the contract.

Remark 4.4.1 *We may think of the contract as a menu of possible payments $\epsilon + \beta W_T^A$, where W_T^A ranges through all possible wealth levels that the Agent can output as he varies his strategies. Most importantly, notice that the Principal cannot infer Agent's actions A by simply looking at W_T^A.*

Because the Principal observes the whole wealth process progressively, we shall actually consider a wider family of contracts of the form:

$$S = \left\{ A \mapsto S(A) := \epsilon(P_{0:T}) + \sum_{t<T} \beta_t \Delta W_{t+1}^A \right\},$$

where $\beta_t \in L^0(\mathcal{F}_t)$ and ϵ is as before, which make better use of her available information. Let us however emphasize that unless the market consists of only one asset, Agent's actions remain unobservable since the most that can be inferred form P and W^A are the affine linear spaces where A must have evolved (see (4.2.1)). However, as in Ou-Yang [2003], we will find that the extra freedom of contracts of type S is not going to be used by the Principal when she restricts herself to incentive-compatible contracts (whereby she proposes actions to the Agent) and thus an optimal contract shall be of the form $\bar{S}$ instead. This is desirable since from a practical point of view it means that the Principal need not waste her time monitoring what happens to the output wealth process at intermediate times, and may be a consequence of our assumption that the Principal does not seek to infer/learn anything about A from observing P and W^A, which we may justify as it being to expensive or time-consuming for her; instead she will recommend the Agent what to do. Regarding notation, we will conveniently refer to a contract as S, (ϵ, β) or $(\epsilon, \{\beta_t\})$ depending on the context.

We will now proceed to motivate the very important definition of incentive compatible contracts, which we formalize in Definition 4.4.1. This definition will be in turn motivated by a set of recursions (that we derive heuristically in a first stage) which also imply that the whole dynamic contracting problem reduces to a sequence of conditional optimization (static) ones. It is the purpose of Theorem 4.4.1 to make all of these steps rigorous.

Given that the Agent decides to follow an investment strategy A during $[0, T]$, which is nothing but an adapted stochastic process ($A_t \in L^0(\mathcal{F}_t)$), his total cost of effort is then $C(A) = \sum_{t=0}^{T-1} c(t, A_t)\Delta t$. For ease of notation we denote $c_t(A) := c(t, A_t)$. By choosing A, the output (wealth) process W is determined by (4.2.1) and hence also the payment of the contract from the Principal to the Agent, amounting to $S(A)$. Viewed from time t this gives the Agent a utility of $U_t^a(S(A) - C(A))$. Using translation invariance we now compute:

$$\begin{aligned} U_t^a\left(S(A)-\sum_t c_t(A_t)\Delta t\right) &= U_t^a\left(\epsilon(P_{0:T})+\sum_{s\geq t}\left\{\beta_s\Delta W_{s+1}^A-c_{s+1}(A_{s+1})\Delta t\right\}\right)\\ &\quad -c_t(A_t)\Delta t+\sum_{s<t}\left\{\beta_s\Delta W_{s+1}^A-c_s(A_s)\Delta t\right\}. \end{aligned} \tag{4.4.1}$$

With translation invariance and time consistency in mind we see that this expression has a recursive structure. This suggests that Agent's optimization problem of finding the best A given a contract $S(\cdot)$ would reduce to the following recursion:

$$\begin{aligned} H_T &= \epsilon(P_{0:T}) \\ H_t &= \operatorname*{ess\,sup}_{A\in L^0(\mathcal{F}_t)^N}\left\{U_t^a\left(H_{t+1}+\beta_t A\Delta\tilde{P}_{t+1}\right)-c_t(A)\Delta t\right\}, \end{aligned} \tag{4.4.2}$$

where we omit for brevity's sake the dependence of H in S. Thus H_t has the interpretation of being the maximal utility the Agent can get, from time t onwards. This intuition will be made rigorous with the aid of Theorem 4.4.1 and its proof. Also notice that adding an $\mathcal{F}_t-$measurable term to ϵ translates additively into H and preserves optimality of any A, down to time t. This shows also that the random variables H_t span all of $L^0(\mathcal{F}_t)$ as the Principal varies the contract parameters.

Since the Principal's utility at time $t=0$ is given by $U_0^p(W_T^A-\epsilon-\sum\beta_s\Delta W_{s+1}^A)$, we see that she will want to steer (simply by modifying ϵ by a constant) H_0 to its lowest admissible level, and in doing so increasing her utility by the same constant (again by translation invariance). Thus, if R denotes Agent's *reservation utility* (i.e. the minimum amount which he needs to be offered at time $t=0$ in order to commit to the contract), we may assume that in Principal's optimization problem the condition $H_0=R$ is binding (i.e. contracts delivering $H_0>R$ cannot be optimal).

Let us now derive a recursion for the Principal. Suppose again that the Agent has chosen A when presented with a contract S, and this time assume that the Principal knows this. Then the derived utility of the Principal as seen from time t is:

$$\begin{aligned} & U_t^p\left(W_T^A-\epsilon-\sum_{s<T}\beta_s\Delta W_{s+1}^A\right)\\ = \;& W_0+\sum_{s<t}(1-\beta_s)A_s\Delta\tilde{P}_{s+1}+U_t^p\left(\sum_{s\geq t}(1-\beta_s)A_s\Delta\tilde{P}_{s+1}-\epsilon\right)\\ = \;& W_0-H_t+\sum_{s<t}(1-\beta_s)A_s\Delta\tilde{P}_{s+1}+U_t^p\left(\sum_{s\geq t}\left[(1-\beta_s)A_s\Delta\tilde{P}_{s+1}-\Delta H_{s+1}\right]\right), \end{aligned}$$

where the identity $\epsilon = H_t + \sum_{s\geq t} \Delta H_{s+1}$ and translation invariance was used. Now using time consistency and translation invariance, we see:

$$\begin{aligned}
& U_t^p\left(\sum_{s\geq t}\left[(1-\beta_s)A_s\Delta\tilde{P}_{s+1} - \Delta H_{s+1}\right]\right) \\
= \quad & U_t^p\left((1-\beta_t)A_t\Delta\tilde{P}_{t+1} + H_t - H_{t+1} + U_{t+1}^p\left(\sum_{s\geq t+1}\left[(1-\beta_s)A_s\Delta\tilde{P}_{s+1} - \Delta H_{s+1}\right]\right)\right) \\
= \quad & U_t^a\left(H_{t+1} + \beta_t A_t\Delta\tilde{P}_{t+1}\right) - c_t(A_t)\Delta t + U_t^p\left(h_{t+1}(A,\beta) + (1-\beta_t)A_t\Delta\tilde{P}_{t+1} - H_{t+1}\right),
\end{aligned}$$

where the definition of H_t was used (as well as the optimality of A), and we defined

$$h_{t+1}(A,\beta) := U_{t+1}^p\left(\sum_{s\geq t+1}\left[(1-\beta_s)A_s\Delta\tilde{P}_{s+1} - \Delta H_{s+1}\right]\right),$$

which represents Principal's future utility from time $t+1$ on. Evidently $h_{t+1}(A,\beta)$ is a short-hand for $h_{t+1}((A_s,\beta_s)_{s\geq t+1})$. With this the preceding identity is more clearly expressed as:

$$h_t(A,\beta) = U_t^a\left(H_{t+1} + \beta_t A_t\Delta\tilde{P}_{t+1}\right) - c_t(A_t)\Delta t + U_t^p\left(h_{t+1}(A,\beta) + (1-\beta_t)A_t\Delta\tilde{P}_{t+1} - H_{t+1}\right). \tag{4.4.3}$$

We are hence suggested to perform the change of variables

$$\Gamma_{t+1} = \beta_t A_t\Delta\tilde{P}_{t+1} + H_{t+1} \in L^0(\mathcal{F}_{t+1}).$$

Let us write $h_t(A,\Gamma)$ whenever in $h_t(A,\beta)$ this substitution has been made. We thus get:

$$h_t(A,\Gamma) = U_t^a(\Gamma_{t+1}) - c_t(A_t)\Delta t + U_t^p\left(h_{t+1}(A,\Gamma) + A_t\Delta\tilde{P}_{t+1} - \Gamma_{t+1}\right). \tag{4.4.4}$$

The following observation is crucial at this point:

Remark 4.4.2 *Because in the previous derivation we assumed that the Principal knows the mappings A_t as functions of $\{P_s\}_{s\leq t}$ for each t, this means that $h_t \in L^0(\mathcal{F}_t)$, since the term $U_t^p(\dots)$ which is a function of $\{P_s, W_s^A\}_{s\leq t}$ becomes then a function of $\{P_s\}_{s\leq t}$ only. That is, whenever the Principal recommends an action and this is implemented by the Agent, all the variables involved in* (4.4.3) *and* (4.4.4) *become price-adapted.*

For a contract we may ask whether a given sequence of efforts recommended by the Principal may be chosen/implemented by the Agent or not. This is the concept behind the following definition of *incentive compatibility*:

Definition 4.4.1 *Given a contract $(\epsilon, \{\beta_t\})$ and a recommended effort $\{A_t\}$, we say that the tuple $(\epsilon, \{\beta_t\}, \{A_t\})$ is incentive compatible if the essential suprema in* (4.4.2) *are attained by A_t for every t.*

Notice that if $(\epsilon, \{\beta_t\}, \{A_t\})$ is incentive compatible, we may construct for every time t the random variable Γ_{t+1}, and quite trivially A_t for every t will also attain

$$\operatorname*{ess\,sup}_a \left[U_t^a\left(\Gamma_{t+1} + \beta_t[a - A_t]\Delta\tilde{P}_{t+1}\right) - c_t(a)\Delta t\right].$$

Therefore we shall also say that $(\{A\}, \{\Gamma\})$ is *incentive compatible* whenever for every t this A_t attains the essential supremum in the previous line.

Let us now define a set which will be relevant in the following:

$$\mathcal{C}_t(\beta) := \left\{ \begin{array}{c} (A,\Gamma) \in [L^0(\mathcal{F}_t)]^N \times L^0(\mathcal{F}_{t+1}) \text{ s.t. for every } \bar{A} \in [L^0(\mathcal{F}_t)]^N : \\ U_t^a(\Gamma) - c_t(A)\Delta t \geq U_t^a\left(\Gamma + \beta[\bar{A} - A]\Delta\tilde{P}_{t+1}\right) - c_t(\bar{A})\Delta t \end{array} \right\}.$$

It is then an easy exercise to check that for an incentive compatible tuple (ϵ, β, A), and after computing the Γ's associated to them, it must hold that $(A_t, \Gamma_{t+1}) \in \mathcal{C}_t(\beta_t)$ for every time t. Therefore at the level of the A's and Γ's the sets $\mathcal{C}_t$ take care of incentive compatibility. In Theorem 4.4.1 we shall work this out in detail.

Upon revisiting equation (4.4.4) for $h(A,\Gamma)$ and keeping in mind the just mentioned interpretation of the sets $\mathcal{C}$, we now introduce a recursion for a process with the interpretation of Principal's future optimal wealth:

$$\begin{aligned} h_T &= 0, \\ h_t &= \operatorname*{ess\,sup}_{\substack{(\beta,A,\Gamma) \\ (A,\Gamma)\in\mathcal{C}_t(\beta)}} U_t^a(\Gamma) - c_t(A)\Delta t + U_t^p\left(h_{t+1} + A\Delta\tilde{P}_{t+1} - \Gamma\right). \end{aligned} \tag{4.4.5}$$

Thus far we have derived heuristically a set of recursions, and motivated the notion of (one-step) incentive compatibility. The next Theorem and its proof show rigorously how the dynamic contracting problem reduces to a sequence of one-step conditional optimization problems.

Theorem 4.4.1 *Assume that the recursions* (4.4.2) *and* (4.4.5) *admit a solution and are attained at each time t. Then Principal's optimal utility at time $t = 0$ equals $W_0 - R + h_0$.*

Further calling $(\beta_t, A_t, \Gamma_{t+1})_{t<T}$ the maximizers attaining h in (4.4.5)*, and defining*

$$\epsilon = \epsilon(P_{0:T}) := R + \sum_{0\leq t<T}\left[\Gamma_{t+1} - \beta_t A_t \Delta\tilde{P}_{t+1} - U_t^a(\Gamma_{t+1}) + c_t(A_t)\Delta t\right],$$

we see that the contract $S = \{(\bar{A}) \mapsto \epsilon(P_{0:T}) + \sum \beta_t \Delta W_{t+1}^{\bar{A}}\}$ is the optimal contract for the Principal, among those satisfying incentive compatibility and reservation utility constraints. The associated optimal effort for the Agent is exactly A and his optimal wealth will be R.

Proof. First we turn our attention to the Agent's recursion. Let $\bar{a}$ be a generic sequence of efforts. From equation (4.4.1), we see that defining

$$H_t(\bar{a}_t, \ldots, \bar{a}_{T-1}) := U_t^a \left(\epsilon(P_{0:T}) + \sum_{s \geq t} \left\{ \beta_s \Delta W_{s+1}^{\bar{a}} - c_{s+1}(\bar{a}_{s+1}) \Delta t \right\} \right) - c_t(\bar{a}_t) \Delta t,$$

we get the recursion

$$\begin{aligned} H_T &= \epsilon(P_{0:T}), \\ H_t(\bar{a}_t, \ldots, \bar{a}_{T-1}) &= U_t^a \left(H_{t+1}(\bar{a}_{t+1}, \ldots, \bar{a}_{T-1}) + \beta_t \bar{a}_t \Delta \tilde{P}_{t+1} \right) - c_t(\bar{a}_t) \Delta t. \end{aligned}$$

Then, in terms of $H_t := \operatorname{ess\,sup}_{a_t, \ldots, a_{T-1}} H_t(a_t, \ldots, a_{T-1})$, we get:

$$H_t(\bar{a}_t, \ldots, \bar{a}_{T-1}) \leq -c(\bar{a}_t) \Delta t + U_t^a \left(\operatorname*{ess\,sup}_{a_{t+1}, \ldots, a_{T-1}} H_{t+1}(a_{t+1}, \ldots, a_{T-1}) + \beta_t \bar{a}_t \Delta \tilde{P}_{t+1} \right).$$

This yields that $H_t \leq \operatorname{ess\,sup}_{a_t} \left\{ -c(a_t) \Delta t + U_t^a \left(H_{t+1} + \beta_t a_t \Delta \tilde{P}_{t+1} \right) \right\}$. For $t = T-1$ this is an equality and by assumption the value H_{T-1} is attained at some $\hat{a}_{T-1}$. Suppose now that equality holds in the previous equation for $t+1, \ldots, T-1$, and H_{t+1} was attained say at $(\hat{a}_{t+1}, \ldots, \hat{a}_{T-1})$. This implies that:

$$\begin{aligned} H_t &\leq \operatorname*{ess\,sup}_{a_t} \left\{ -c(a_t) \Delta t + U_t^a \left(H_{t+1}(\hat{a}_{t+1}, \ldots, \hat{a}_{T-1}) + \beta_t a_t \Delta \tilde{P}_{t+1} \right) \right\} \\ &\leq \operatorname*{ess\,sup}_{a_t, \ldots, a_{T-1}} \left\{ -c(a_t) \Delta t + U_t^a \left(H_{t+1}(a_{t+1}, \ldots, a_{T-1}) + \beta_t a_t \Delta \tilde{P}_{t+1} \right) \right\} \\ &= \operatorname*{ess\,sup}_{A_t, \ldots, A_{T-1}} H_t(A_t, \ldots, A_{T-1}) \\ &=: H_t. \end{aligned}$$

So indeed at time t also $H_t = \operatorname{ess\,sup}_{a_t} \left\{ -c(a_t) \Delta t + U_t^a \left(H_{t+1} + \beta_t a_t \Delta \tilde{P}_{t+1} \right) \right\}$ holds and by assumption the last term is attained at some $\hat{a}_t$, from which H_t is attained at $(\hat{a}_t, \ldots, \hat{a}_{T-1})$. This closes the inductive step, and therefore the desired recursion holds. Now we will establish rigorously recursion (4.4.4) (equivalently (4.4.3)). To this end we denote by $\beta = (\beta_t)_t$ a generic decision variable for the Principal and $a = (a_t)_t$ where $a_t \in L^0(\mathcal{F}_t)^N$, a corresponding optimal effort for the Agent. Let

$$N := \sum_{s \geq t+1} \left[(1 - \beta_s) a_s \Delta \tilde{P}_{s+1} - \Delta H_{s+1} \right].$$

Then using the expression for H_t given in (4.4.2), and denoting $\Gamma = \beta_t a_t \Delta \tilde{P}_{t+1} + H_{t+1}$, we get:

$$\begin{aligned} U_t^p \left(\sum_{s \geq t} \left[(1 - \beta_s) a_s \Delta \tilde{P}_{s+1} - \Delta H_{s+1} \right] \right) &= U_t^p \Big((1 - \beta_t) a_t \Delta \tilde{P}_{t+1} - H_{t+1} - c_t(a_t) \Delta t + N \\ &\qquad + U_t^a \left(H_{t+1} + \beta_t a_t \Delta \tilde{P}_{t+1} \right) \Big) \end{aligned}$$

$$
\begin{aligned}
&= U_t^p\left(a_t\Delta\tilde{P}_{t+1} - \Gamma + U_t^a(\Gamma) - c_t(a_t)\Delta t + N\right)\\
&= U_t^a(\Gamma) - c_t(a_t)\Delta t + U_t^p\left(a_t\Delta\tilde{P}_{t+1} - \Gamma + N\right).
\end{aligned}
$$

And now, applying time-consistency and translation invariance in the last term above we get:

$$
U_t^p\left(\sum_{s\geq t}\left[(1-\beta_s)a_s\Delta\tilde{P}_{s+1} - \Delta H_{s+1}\right]\right) = U_t^a(\Gamma) - c_t(a_t)\Delta t + U_t^p\left(a_t\Delta\tilde{P}_{t+1} - \Gamma + U_{t+1}^p(N)\right).
$$

Therefore calling

$$
h_{t+1}(a,\Gamma) = U_{t+1}^p(N),
$$

we obtain recursion (4.4.4). That is to say, if (a,Γ) does not satisfy this recursion, they will not be chosen by the Principal. In the same way we conclude for (a,β) and recursion (4.4.3). With these recursions for $h_t(\cdot)$ already established, we proceed to prove (4.4.5). First recall that actually $h_t(a,\Gamma)$ is a short-hand for $h_t((a_s,\Gamma_{s+1})_{s\geq t})$. From this and (4.4.4) we have:

$$
h_t((\bar{a}_s,\bar{\Gamma}_{s+1})_{s\geq t}) \leq U_t^p\left(\operatorname*{ess\,sup}_{A,\Gamma} h_{t+1}(A,\Gamma) + \bar{a}_t\Delta\tilde{P}_{t+1} - \bar{\Gamma}_{t+1}\right) + U_t^a(\bar{\Gamma}_{t+1}) - c_t(\bar{a}_t)\Delta t.
$$

This yields

$$
\begin{aligned}
h_t \quad &:= \operatorname*{ess\,sup}_{\substack{(a_s,\Gamma_{s+1})\in\mathcal{C}_s(\beta_s)\\ s\geq t}} h_t\left((a_s,\Gamma_{s+1})_{s\geq t}\right)\\
&\leq \operatorname*{ess\,sup}_{(a_t,\Gamma_{t+1})\in\mathcal{C}_t(\beta_t)} U_t^p\left(h_{t+1} + a_t\Delta\tilde{P}_{t+1} - \Gamma_{t+1}\right) + U_t^a(\Gamma_{t+1}) - c_t(a_t)\Delta t.
\end{aligned}
$$

For $t = T-1$ this is an equality (we define $h_T = 0$) and by assumption the value h_{T-1} is attained, say at $(\hat{a}_{T-1},\hat{\Gamma}_T)$. Now suppose that for $t+1,\ldots,T-1$ equality holds in the previous equation, and h_{t+1} was attained say at $(\hat{a}_s,\hat{\Gamma}_{s+1})_{s\geq t+1}$. This implies, thanks to (4.4.4), that:

$$
\begin{aligned}
h_t \quad &\leq \operatorname*{ess\,sup}_{(a_t,\Gamma_{t+1})\in\mathcal{C}_t(\beta_t)} U_t^p\left(h_{t+1}(\hat{a},\hat{\Gamma}) + a_t\Delta\tilde{P}_{t+1} - \Gamma_{t+1}\right) + U_t^a(\Gamma_{t+1}) - c_t(a_t)\Delta t\\
&\leq \operatorname*{ess\,sup}_{\substack{(a_s,\Gamma_{s+1})\in\mathcal{C}_s(\beta_s)\\ s\geq t}} U_t^p\left(h_{t+1}(a,\Gamma) + a_t\Delta\tilde{P}_{t+1} - \Gamma_{t+1}\right) + U_t^a(\Gamma_{t+1}) - c_t(a_t)\Delta t\\
&= \operatorname*{ess\,sup}_{\substack{(a_s,\Gamma_{s+1})\in\mathcal{C}_s(\beta_s)\\ s\geq t}} h_t\left((a_s,\Gamma_{s+1})_{s\geq t}\right)\\
&=: \; h_t.
\end{aligned}
$$

So indeed equality holds also at time t and by assumption the last term is attained at some $(\hat{a}_t,\hat{\Gamma}_{t+1})$, from which h_t is attained at $(\hat{a}_s,\hat{\Gamma}_{s+1})_{s\geq t}$. This closes the inductive step and therefore the desired recursion holds.

The validity of the change of variables $\beta_t a_t \Delta\tilde{P}_{t+1}+H_{t+1} \to \Gamma_{t+1}$ and the introduction of $\mathcal{C}(\beta)$ as a constraint inducing incentive compatibility are now obvious. This means that h represents the future wealth prospects of the Principal. Hence at time $t=0$ we obtain a solution for the whole Principal's problem, proving as well that Principal's optimal wealth is $W_0 - R + h_0$.

We proceed now to prove that a solution to Principal's recursion delivers indeed an optimal (dynamic) contract, and that the Agent behaves as predicted. Call $(\beta_t, A_t, \Gamma_{t+1})_t$ the optimal quantities attaining h in (4.4.5). Define ϵ and the contract S as in the statement of the present theorem. Then:

$$\begin{aligned} U^a_{T-1}\left(\epsilon + \beta_{T-1}a_{T-1}\Delta\tilde{P}_T\right) - c_{T-1}(a_{T-1})\Delta t \quad &= \sum_{0\le t<T-1}\left[\Gamma_{t+1} - \beta_t A_t \Delta\tilde{P}_{t+1} - U^a_t(\Gamma_{t+1}) + c_t(A_t)\Delta t\right] \\ &\quad R + [c_{T-1}(A_{T-1}) - c_{T-1}(a_{T-1})]\Delta t - U^a_{T-1}(\Gamma_T) \\ &\quad + U^a_{T-1}\left(\Gamma_T - \beta_{T-1}A_{T-1}\Delta\tilde{P}_T + \beta_{T-1}a_{T-1}\Delta\tilde{P}_T\right). \end{aligned}$$

By definition of $\mathcal{C}(\beta)$ the sum of the last terms is smaller or equal than 0, and exactly zero when $a_{T-1} = A_{T-1}$. Therefore

$$\begin{aligned} \operatorname*{ess\,sup}_{a_{T-1}}\left\{U^a_{T-1}\left(\epsilon + \beta_{T-1}a_{T-1}\Delta\tilde{P}_T\right) - c_{T-1}(a_{T-1})\Delta t\right\} \quad &= \quad R + \sum_{0\le t<T-1}[\Gamma_{t+1} - \beta_t A_t \Delta\tilde{P}_{t+1} \\ &\quad + c_t(A_t)\Delta t - U^a_t(\Gamma_{t+1})]. \end{aligned}$$

This shows that at time $T-1$ the Agent chooses A_{T-1} when presented with (ϵ,β). If we define $H_T = \epsilon$, we are thus entitled to call H_T the value (left hand side or right one) in the above equality. Now inductively, assume that for $t+1,\dots,T-1$ it holds that

$$\begin{aligned} \operatorname*{ess\,sup}_{a_{t+1}}\left\{U^a_{t+1}\left(H_{t+2} + \beta_{t+1}a_{t+1}\Delta\tilde{P}_{t+2}\right) - c_{t+1}(a_{t+1})\Delta t\right\} \quad &= \quad R + \sum_{0\le s<t+1}[\Gamma_{s+1} - \beta_s A_s \Delta\tilde{P}_{s+1} \\ &\quad + c_s(A_s)\Delta t - U^a_s(\Gamma_{s+1})], \end{aligned}$$

where the above value (right hand side or the left one) we call H_{t+1}, and we assume that it is attained at (A_{t+1}). Now, we have

$$\begin{aligned} U^a_t\left(H_{t+1} + \beta_t a_t \Delta\tilde{P}_{t+1}\right) - c_t(a_t)\Delta t \quad &= \sum_{0\le s<t}\left[\Gamma_{s+1} - \beta_s A_s \Delta\tilde{P}_{s+1} - U^a_s(\Gamma_{s+1}) + c_s(A_s)\Delta t\right] \\ &\quad + c_t(A_t)\Delta t + U^a_t\left(\Gamma_{t+1} - \beta_t A_t\Delta\tilde{P}_{t+1} + \beta_t a_t \Delta\tilde{P}_{t+1}\right) \\ &\quad - c_t(a_t)\Delta t - U^a_t(\Gamma_{t+1}) + R, \end{aligned}$$

and we see again that by definition of $\mathcal{C}(\beta)$ the sum of the last term is at most 0 and equal to this number for $a_t = A_t$. Therefore, also at time t the Agent chooses A_t when presented with the contract (ϵ,β). Again, we are entitled to write H_t for the common value:

$$\begin{aligned} \operatorname*{ess\,sup}_{a_t}\left\{U^a_t\left(H_{t+1} + \beta_t a_t \Delta\tilde{P}_{t+1}\right) - c_t(a_t)\Delta t\right\} \quad &= \quad R + \sum_{0\le s<t}\left[\Gamma_{s+1} - \beta_s A_s \Delta\tilde{P}_{s+1}\right. \\ &\quad \left. - U^a_s(\Gamma_{s+1}) + c_s(A_s)\Delta t\right]. \end{aligned}$$

By induction we have proven that the contract S (defined from (ϵ, β)) is optimal for the Principal and incentive compatible (notice that automatically $H_0 = R$), and the Agent indeed chooses A under this contract. This finishes the proof. ■

Remark 4.4.3 *The same reasoning would have been applicable if either Principal or Agent had had a random initial endowment (call them E^p and E^a respectively), or for that matter, had had to hedge some claims. The only difference is that in recursion* (4.4.2) *the ending-condition has to be changed to $H_T = \epsilon(P_{0:T}) + E^a$, and for the Principal, the ending condition in recursion* (4.4.5) *has to be changed to $h_T = E^p$. All of the attainability results in the next sections would still apply, but the catch is that several (conditional) integrability conditions need be satisfied and their fulfilment will obviously depend on the specific random endowments.*

We define now an auxiliary unconstrained version of (4.4.5) and prove that if such a problem is well-posed, it yields a solution to the original one-step problem, and the corresponding $\beta = 1$ is optimal. If this is the case for every time, this implies that the first-best solution is implementable if it exists. The importance of this simple result is that we may dispense with the non-convex sets $\mathcal{C}_t$, thus turning a non-concave maximization problem into a concave one.

Theorem 4.4.2 *Assume that the following problem is finite and attainable:*

$$\Sigma := \operatorname*{ess\,sup}_{(A,\Gamma)\in[L^0(\mathcal{F}_t)]^N\times L^0(\mathcal{F}_{t+1})} U_t^a(\Gamma) - c_t(A)\Delta t + U_t^p\left(h_{t+1} + A\Delta\tilde{P}_{t+1} - \Gamma\right). \tag{4.4.6}$$

Call $\left(\hat{A}, \hat{\Gamma}\right)$ any maximizer. Then $\left(\hat{A}, \hat{\Gamma}\right) \in \mathcal{C}_t(1)$ and therefore

$$\Sigma \quad = \quad \operatorname*{ess\,sup}_{\substack{(\beta,A,\Gamma)\\(A,\Gamma)\in\mathcal{C}_t(\beta)}} U_t^a(\Gamma) - c_t(A)\Delta t + U_t^p\left(h + A\Delta\tilde{P}_{t+1} - \Gamma\right).$$

Proof. Let $\left(\hat{A}, \hat{\Gamma}\right)$ be a maximizer for (4.4.6). For arbitrary A, define $\Gamma = \hat{\Gamma} + (A - \hat{A})\Delta\tilde{P}$. Plugging in that $\left(\hat{A}, \hat{\Gamma}\right)$ is better than (A, Γ) for (4.4.6), we see that the terms involving U^p cancel out and we are left with:

$$U_t^a(\hat{\Gamma}) - c_t(\hat{A})\Delta t \geq U_t^a(\hat{\Gamma} + (A - \hat{A})\Delta\tilde{P}) - c_t(A)\Delta t, \tag{4.4.7}$$

and this holds for every $A \in [L^0(\mathcal{F}_t)]^N$. This means that $\left(\hat{A}, \hat{\Gamma}\right) \in \mathcal{C}_t(1)$, further implying that the value of the constrained and the unconstrained problems coincide.

■

Remark 4.4.4 *Since we are dealing with recommended efforts and incentive compatible contracts, we have that $U^p(\dots)$ only depends on P. It is this function of P that appears in Σ in the above result. We thus may and will simply consider the restriction of U_t^p to $L^0(\mathcal{F}_T)$, recalling that it then satisfies the usual hypotheses w.r.t. $\mathcal{F}_t$ as well.*

The previous proof heavily relies on the fact that contracts are linear. Indeed by varying $\hat{\Gamma}$ in any direction of the form $(A - \hat{A})\Delta\tilde{P}$ and from linearity of contracts the term in the objective function involving Principal's utility became irrelevant, making it possible to compare the values of Agent's utilities.

Remark 4.4.5 *A contract $\bar{S}$ for which the β's equal 1 has the form S described at the beginning of this section. Furthermore, in such case the interpretation is that the Agent keeps the output W_T to himself and pays in return the derivative $-\epsilon$ to the Principal.*

Because the unrestricted problem gives a solution to the original one, provided the former is attained, we shall turn in section 4.6 our attention to the question of attainability of the unconstrained problems. Before that we need to introduce the technical language and results that permit to study the attainability issue.

4.5 Digression into conditional analysis essentials

Let us present some results concerning finite-dimensional conditional analysis, as well as conditional analysis on L^p spaces. This sections is mainly a survey, although suitable extensions of known results are proved when needed.

Finite-dimensional case

We follow Cheridito et al. [b]. On our probability space $(\Omega, \mathcal{F}, \mathbb{P})$ call L the set of all measurable functions (possibly infinite-valued) and L^0 the finite ones. As usual consider almost-sure identification and ordering on this set. Further call $\overline{L} = \{X \in L : X > -\infty\}$ (and $\underline{L} = \{X \in L : X < \infty\}$) and $\mathbb{N}(\mathcal{F})$ the set of variables in L^0 taking values in $\mathbb{N}$.

The spaces to be surveyed here are the finite cartesian products of L^0 spaces. From now on we fix $N \in \mathbb{N}$ and call $E = [L^0(\mathcal{F})]^N$. We view now this space as a finite-dimensional topological $L^0(\mathcal{F})$-module over the ring $L^0(\mathcal{F})$. On E we define the *conditional norm* $\|X\| = (X \cdot X)^{\frac{1}{2}}$ (notice that this is a random variable), where the product is the euclidean one.

Definition 4.5.1
A set $C \subset E$ is called:

- *stable if $\mathbb{1}_A X + \mathbb{1}_{A^c} Y \in C$, for every $X, Y \in C$, $A \in \mathcal{F}$*
- *$\sigma-$stable if $\sum_{n\in\mathbb{N}} \mathbb{1}_{A_n} X_n \in C$, for every sequence $(X_n) \subset C$ and every partition $(A_n) \subset \mathcal{F}$ of Ω*
- *L^0-convex if $\lambda X + (1-\lambda) Y \in C$, for every $X, Y \in C$ and $\lambda \in L^0$ with values in $[0,1]$*
- *sequentially closed if it contains all the limits of its a.s. converging sequences.*
- *L^0-bounded if $\operatorname{ess\,sup}_{X\in C} \|X\| \in L^0$.*

Notice that a stable and sequentially closed set is automatically $\sigma-$stable. We define for $M \in \mathbb{N}(\mathcal{F})$ and $(X_n) \subset E$ the element $X_M = \sum_{n\in\mathbb{N}} \mathbb{1}_{M=n} X_n \in E$. Notice that if the former sequence belongs to a $\sigma-$stable set, then the latter element too.

The following result is a generalization of the classical Bolzano-Weierstrass Theorem:

Theorem 4.5.1 *Let $(X_n) \subset E$ be L^0-bounded. Then there exists $X \in E$ and a sequence $(N_n) \in \mathbb{N}(\mathcal{F})$ such that $N_{n+1} > N_n$ and $X = \lim_{n\to\infty} X_{N_n}$ a.s.*
Also, let $(x_n) \subset L^0$ be such that $x := \limsup x_n \in L^0$. Then there exists a sequence $(N_n) \in \mathbb{N}(\mathcal{F})$ such that $N_{n+1} > N_n$ and $x = \lim_{n\to\infty} x_{N_n}$ a.s.

Proof. For the first statement refer to [Cheridito et al., b, Theorem 3.8]. For the second, define $N_0 = 0$ and $N_n = \min\{m > N_{n-1} : x_m \geq x - 1/n\}$. Then clearly $N_n \in \mathbb{N}(\mathcal{F})$ and $N_{n+1} > N_n$, from which $N_n \geq n$ follows. Now, notice $\sup_{m\geq n} x_m \geq \sup_{m\geq N_n} x_m \geq x_{N_n} \geq x - 1/n$ a.s., from which $x = \lim_{n\to\infty} x_{N_n}$ a.s. ∎

Clearly if the sequences above had belonged to a stable, sequentially closed and L^0-bounded set, then also the randomized subsequences as well as its limits would have belonged to the same set.

As in the euclidean case, convexity opens the way to a necessary and sufficient characterization of boundedness (see [Cheridito et al., b, Theorem 3.13]):

Theorem 4.5.2 *Let C be a sequentially closed L^0-convex subset of E which contains 0. Then C is L^0-bounded if and only if for any $X \in C\backslash\{0\}$ there exists a $k \in \mathbb{N}$ such that $kX \notin C$.*

Definition 4.5.2
Let $C \subset E$. A function $f : C \to L$ is called:

- *L^0-lower semicontinuous at $X \in C$ if $f(X) \leq \liminf f(X_n)$ for every sequence $(X_n) \subset C$ with a.s. limit X.*
- *L^0-continuous at $X \in C$ if $f(X) = \lim f(X_n)$ for every sequence $(X_n) \subset C$ with a.s. limit X.*
- *L^0-convex if $f(\lambda X + (1-\lambda)Y) \leq \lambda f(X) + (1-\lambda) f(Y)$, for every $X, Y \in C$ and $\lambda \in L^0$ with values in $[0, 1]$*
- *stable if $f(\mathbb{1}_A X + \mathbb{1}_{A^c} Y) = \mathbb{1}_A f(X) + \mathbb{1}_{A^c} f(Y)$, for every $X, Y \in C$, $A \in \mathcal{F}$.*

Of course stability, as defined, holds equivalently for any finite partition. Also it is assumed that for the last two points the set C must be itself L^0-convex or stable respectively. The dual definitions of upper semicontinuity and L^0-concavity are the obvious ones. Strict L^0-convexity is defined with an $<$, on the set $\{X \neq \lambda X + (1-\lambda)Y \neq Y\}$. Finally the function is called (upper/lower semi)continuous on C if it is so on every point of C. Notice that if f is continuous and stable over a $\sigma-$stable and sequentially closed set, then it satisfies the stability property for countable partitions too. What is more, if f is only L^0-convex or L^0-concave, then it is automatically local (meaning $\mathbb{1}_A f(X) = \mathbb{1}_A f(Y)$ whenever $\mathbb{1}_A X = \mathbb{1}_A Y$), which in itself directly implies that it also satisfies the stability property for countable partitions.

Remark 4.5.1 *By the previous discussion, if f is a $L^0(\mathcal{F})-$convex (resp. -concave) function then the set $\{f \leq s\}$ (resp. $\{f \geq s\}$) is $L^0(\mathcal{F})-$convex and $\sigma-$stable, for every $s \in L^0(\mathcal{F})$.*

The previous remark will sometimes be used along the following result:

Lemma 4.5.1 *If a non-empty set $C \subset E$ is $\sigma-$stable and is not L^0-bounded, then there is a set $\tilde{\Omega}$ with $\mathbb{P}(\tilde{\Omega}) > 0$ and a sequence $\{X_n\} \subset C$ such that, for every $n \in \mathbb{N}$, $|X_n| \geq n$ over $\tilde{\Omega}$*

Proof. Define $U_n := \{B \in \mathcal{F} : \exists X \in C, |X| \geq n \text{ on } B\}$, which is non-empty by unboundedness of C. Next define $A_n := \left\{ \operatorname*{ess\,sup}_{B \in U_n} \mathbb{1}_B = 1 \right\}$, which is a measurable way of defining $\bigcup_{B \in U_n} B$. Noticing that the A_n are decreasing let us introduce $A := \bigcap_n A_n$. Assuming that $\mathbb{P}(A) = 0$, or equivalently that $\mathbb{P}\left(\cup_n A_n^c\right) = 1$, then for every $X \in C$

$$|X| = \left| \sum_n X \mathbb{1}_{\{A_n^c \cap A_{n-1}\}} \right| \leq \sum_n |X| \mathbb{1}_{\{A_n^c \cap A_{n-1}\}} \leq \sum_n n \mathbb{1}_{\{A_n^c \cap A_{n-1}\}} \in L^0(\mathcal{F}).$$

Since X was arbitrary in C, this implies that C is $L^0(\mathcal{F})-$bounded. Therefore $\mathbb{P}(A) > 0$ must hold.

By definition of ess sup we have that there exist $\{B^{l,n}\}_l \in U_n$ such that $\operatorname*{ess\,sup}_{B \in U_n} \mathbb{1}_B = \sup_l \mathbb{1}_{B^{l,n}}$ a.s. Therefore also $A_n = \bigcup_l B^{l,n}$ a.s. Taking $X^{l,n}$ such that $|X^{l,n}| \geq n$ on $B^{l,n}$, and fixing an $X^* \in C$ arbitrary, let us define:

$$X^{(n)} := X^* \mathbb{1}_{\{(\bigcup_l B^{l,n})^c\}} + \sum_l X^{l,n} \mathbb{1}_{\{B^{l,n} \cap (\cup_{m<l} B^{m,n})^c\}} + X^{0,n} \mathbb{1}_{B^{0,n}},$$

which belongs to C thanks to $\sigma-$stability. Clearly

$$|X^{(n)}| \geq n \mathbb{1}_{\{\bigcup_l B^{l,n}\}} + |X^*| \mathbb{1}_{\{(\bigcup_l B^{l,n})^c\}},$$

and therefore a.s. $|\mathbb{1}_A X^{(n)}| \geq n \mathbb{1}_A$. Thus we have that $|X^{(n)}| \geq n$ on A for every n. ∎

Remark 4.5.2 *The previous lemma is actually implied by a part of the proof of [Cheridito et al., b, Theorem 4.13], since all the authors use is $\sigma-$stability of the set under consideration (which is implied by their stronger assumptions of conditional convexity and sequential closedness). We chose to give a self-contained proof here.*

With the applications in mind, the following conditional optimization theorem will be useful (see [Cheridito et al., b, Theorem 4.4]):

Theorem 4.5.3 *Let C be a sequentially closed and stable subset of E and $f : C \to \overline{L}$ a L^0-lower semicontinuous, stable function. Assume there exists an $X_0 \in C$ such that the set $\{X \in C : f(X) \leq f(X_0)\}$ is L^0-bounded. Then there exists an $\hat{X} \in C$ such that*

$$f\left(\hat{X}\right) = \operatorname*{ess\,inf}_{X\in C} f(X)$$

If f and C are L^0-convex then the "argmin" set is also L^0-convex, and if in addition f is strictly L^0-convex then $\hat{X}$ is the sole (a.s.) optimizer.

A last concept that will prove useful later on is that of orthogonal complement. For a non-empty set $C \subset E$ its *orthogonal complement* is given by $C^{\perp} := \{X \in E : X \cdot Y = 0, \forall Y \in C\}$. In case C were a $\sigma-$stable L^0-linear set, then $C + C^{\perp} = E$ and $C \cap C^{\perp} = \{0\}$ (see Cheridito et al. [b]).

In the next section, a very specific instance of infinite-dimensional analysis will be briefly introduced.

Conditional analysis on L^p spaces

Let $\mathcal{F}$ be a sub sigma-algebra of $\mathcal{G}$. The conditional version of the L^p spaces will be introduced. For a thorough treatment one may see Filipović et al. [2012].

For every $p \in [1, +\infty]$ define:

$$||X||_p \quad = \quad \begin{cases} \mathbb{E}[|X|^p|\mathcal{F}] & \text{if } p \in [1, +\infty) \\ \operatorname{ess\,inf}\{Y \in L^0_+(\mathcal{F}) \text{ s.t. } Y \geq |X|\} & \text{if } p = +\infty. \end{cases}$$

This is well defined for every $X \in L^0(\mathcal{G})$. We further define the conditional L^p-space $L^p_{\mathcal{F}}(\mathcal{G}) := \{X \in L^0(\mathcal{G}) \text{ st. } ||X||_p \in L^0(\mathcal{F})\}$. As remarked in Filipović et al. [2012], $L^p_{\mathcal{F}}(\mathcal{G})$ is a topological $L^0(\mathcal{F})-$module over the topological ring $L^0(\mathcal{F})$, and $||\cdot||_p$ is an $L^0(\mathcal{F})-$norm inducing the module topology on $L^p_{\mathcal{F}}(\mathcal{G})$.

A function $U : L^p_{\mathcal{F}}(\mathcal{G}) \to \underline{L}^0$ will be called:

- $L^0(\mathcal{F})-$Concave: if $U(\lambda X + (1-\lambda)X') \geq \lambda U(X) + (1-\lambda)U(X')$ for every $\lambda \in L^0(\mathcal{F}) \cap [0,1]$ and every $X, X' \in L^p_{\mathcal{F}}(\mathcal{G})$
- Proper: if $\exists X \in L^p_{\mathcal{F}}(\mathcal{G})$ such that $U(X) > -\infty$ and $\forall X' \in L^p_{\mathcal{F}}(\mathcal{G})$ it holds $U(X) < \infty$
- $L^p_{\mathcal{F}}(\mathcal{G})$-upper semicontinuous: if for every net $\{X_\alpha\} \subset L^p_{\mathcal{F}}(\mathcal{G})$ converging to some X in conditional norm, it holds that $\operatorname{ess\,inf}_\beta \operatorname{ess\,sup}_{\alpha\geq\beta} U(X_\alpha) \leq U(X)$
- Monotone: if $U(X) \geq U(X')$ whenever $X \geq X'$
- Translation invariant: if $U(X+Y) = U(X) + Y$ for every $X \in L^p_{\mathcal{F}}(\mathcal{G})$ and $Y \in L^0(\mathcal{F})$

The whole purpose of the present section is the following representation result of a functional satisfying the above properties. This is a re-phrasing of [Filipović et al., 2012, Corollary 3.14]:

Theorem 4.5.4 *Let $U : L^p_{\mathcal{F}}(\mathcal{G}) \to \underline{L}^0(\mathcal{F})$ be proper, $L^p_{\mathcal{F}}(\mathcal{G})$-upper semicontinuous, monotone, translation invariant and $L^0(\mathcal{F})-$concave. Define q as the Hölder conjugate of p, $\mathcal{W} = \{Z \in L^q_{\mathcal{F}}(\mathcal{G}) : Z \geq 0, \mathbb{E}[Z|\mathcal{F}] = 1\}$ and call $\alpha(Z) = \operatorname*{ess\,sup}_{X \in L^p_{\mathcal{F}}(\mathcal{G})} \{U(X) - \mathbb{E}[ZX|\mathcal{F}]\}$. Then:*

$$U(X) = \operatorname*{ess\,inf}_{Z \in \mathcal{W}} \{\mathbb{E}[ZX|\mathcal{F}] + \alpha(Z)\}$$

In the next Proposition we prove how $L^p_{\mathcal{F}}(\mathcal{G})-$upper semicontinuity is a consequence of L^0-upper semicontinuity (see section 4.3), which is nothing but a sequential and almost-sure version of upper semicontinuity. This result will be invoked in Remark 4.6.3.

Proposition 4.5.1 *Let $U : L^p_{\mathcal{F}}(\mathcal{G}) \to \underline{L}^0(\mathcal{F})$ be L^0-upper semicontinuous (i.e. sequentially, almost surely). Then U is also $L^p_{\mathcal{F}}(\mathcal{G})-$upper semicontinuous.*

Proof. By [Filipović et al., 2009, Lemma 3.10], it is enough to prove that the sets $K_k := \{X \in L^p_{\mathcal{F}}(\mathcal{G}) : U(X) \geq k\}$ are conditionally closed for every $k \in L^0(\mathcal{F})$. We will prove that their complements are conditionally open. So let us fix such a k and by contradiction, suppose that $(K_k)^c$ is not open. We thus take X such that $U(X) < k$ on a non-negligible set (i.e. $X \in (K_k)^c$) and such that for every $N \in \mathbb{N}(\mathcal{F})$ we have that $K_k \cap B(X, 1/N) \neq \emptyset$, where $B(X, 1/N) = \{Z : \mathbb{E}(|Z - X|^p|\mathcal{F}) \leq 1/N\}$. This means that we can find, for every $N \in \mathbb{N}(\mathcal{F})$, an element $X_N \in B(X, 1/N)$ such that $U(X_N) \geq k$ a.s. A trivial adaptation of Markov's inequality gives that for every $\epsilon \in L^0(\mathcal{F})_{++}$, $\mathbb{P}(|X_N - X| \geq \epsilon|\mathcal{F}) \leq \frac{\mathbb{E}(|X_N - X|^p|\mathcal{F})}{\epsilon^p}$. From this we may find for every natural number n an element $M_n \in \mathbb{N}(\mathcal{F})$ such that:

- For every $N \in \mathbb{N}(\mathcal{F})$ st. $N \geq M_n$ holds that $\mathbb{P}(|X_N - X| \geq 1/n|\mathcal{F}) \leq 1/n^2$ a.s.
- For every n: $M_{n+1} > M_n$ a.s.

Now, we will use a "Borel-Cantelli Lemma"-type reasoning in order to prove that this sequence $\{X_{M_n}\}$ converges almost surely to X. First notice that for a fixed $l \in \mathbb{N}$:

$$\sum_{n \in \mathbb{N}} \mathbb{P}(|X_{M_n} - X| \geq 1/l|\mathcal{F}) \leq \sum_{n \leq l} \mathbb{P}(|X_{M_n} - X| \geq 1/l|\mathcal{F}) + \sum_{n > l} \mathbb{P}(|X_{M_n} - X| \geq 1/n|\mathcal{F}),$$

and since the last term is bounded above by $\sum_{n>l} 1/n^2$ we see that the original sum belongs to $L^0(\mathcal{F})$ (and so is a.s. finite). Define now $i.o. \{|X_{M_\cdot} - X| \geq 1/l\} := \bigcap_{m \in \mathbb{N}} \bigcup_{n \geq m} \{|X_{M_n} - X| \geq 1/l\}$. Then:

$$\begin{aligned} \mathbb{P}\left(i.o. \{|X_{M_\cdot} - X| \geq 1/l\} |\mathcal{F}\right) &\leq \mathbb{P}\left(\bigcup_{n \geq m} \{|X_{M_n} - X| \geq 1/l\}|\mathcal{F}\right) \\ &\leq \sum_{n \geq m} \mathbb{P}(|X_{M_n} - X| \geq 1/l|\mathcal{F}), \end{aligned}$$

and so the left hand-side does not depend on m whereas the right one tends a.s. to 0 as m increases. This shows that $\mathbb{P}\left(i.o. \left\{|X_{M.} - X| \geq 1/l\right\} | \mathcal{F}\right) = 0$ almost surely, and taking expectations $\mathbb{P}\left(i.o. \left\{|X_{M.} - X| \geq 1/l\right\}\right) = 0$. Since this holds for every l a natural number, we conclude that indeed $\{X_{M_n}\}$ converges almost surely to X.

Finally we have by assumption that $k \leq \limsup_n U(X_{M_n}) \leq U(X)$ a.s., but also $U(X) < k$ on a non-negligible set, which is a contradiction. This completes the proof. ■

Let us introduce a concept which will be useful in the proof of Theorem 4.6.2. It is a slight extension of [Delbaen and Schachermayer, 1994, Lemma A1.1], which is in itself a Komlos-type result:

Lemma 4.5.2 *Let $\{\xi_n\}_n$ be $[0, +\infty)$-valued random variables defined on a common probability space $(\Omega, \mathcal{G}, \mathbb{P})$, take $\mathcal{F}$ a sub-sigma algebra and assume $C := conv\{\xi_n : n \in \mathbb{N}\}$ is "$\mathcal{F}$-conditionally bounded in probability," defined as:*

$$\forall \epsilon \in L^0_+(\mathcal{F}), \exists a \in L^0(\mathcal{F}) \text{ such that } \forall h \in C, \mathbb{P}(h \geq a|\mathcal{F}) \leq \epsilon.$$

Then there exists an $[0, +\infty)$-valued random variable X and a sequence $\{x_n\}_n$, where x_n belongs to the convex hull of $\{\xi_n, \xi_{n+1}, \dots\}$ such that $x_n \to X$ almost surely.

Proof. Following the proof in [Delbaen and Schachermayer, 1994, Lemma A1.1], the sequence $\{x_n\}_n$ with the desired convergence is obtained. It remains to be shown that its limit X is a.s. finite. But notice that by conditional Fatou's lemma $\mathbb{P}(X \geq a|\mathcal{F}) \leq \liminf \mathbb{P}(x_n \geq a|\mathcal{F})$, for every $a \in L^0(\mathcal{F})$. Therefore, fixing $\epsilon \in L^0_+(\mathcal{F})$ the existence of $a \in L^0(\mathcal{F})$ such that $\mathbb{P}(x_n \geq a|\mathcal{F}) \leq \epsilon$ shows that $\mathbb{P}(X \geq a|\mathcal{F}) \leq \epsilon$. On the other hand, $\mathbb{P}(X = \infty|\mathcal{F}) \leq \mathbb{P}(X \geq a|\mathcal{F}) \leq \epsilon$. Since ϵ was arbitrary this shows that $\mathbb{P}(X = \infty|\mathcal{F}) = 0$ a.s. and therefore $\mathbb{P}(X = \infty) = 0$. ■

Notice that we have referred to several types of boundedness so far. A set C of $L^0(\mathcal{F})$ was called $L^0(\mathcal{F})$-bounded when $\operatorname{ess\,sup}_{c \in C} |c| < \infty$ (this is pointwise boundedness, more or less). In Delbaen and Schachermayer [1994] they use another notion of boundedness, which we may call "in probability": $\forall \epsilon \in \mathbb{R}_{++}, \exists a \in \mathbb{R}_+$ such that $\mathbb{P}(|c| \geq a) \leq \epsilon$ for every $c \in C$. The latter notion is weaker than the former, or than norm boundedness in some L^p space. Finally, we introduced "$\mathcal{F}$-conditionally bounded in probability." It is clear that if the set C (now assumed to consist of functions measurable in a bigger sigma-algebra $\mathcal{G}$) is bounded in some conditional $L^p_{\mathcal{F}}(\mathcal{G})$ space then it is also bounded $\mathcal{F}-$conditionally in probability (for instance, if $\exists f \in L^0(\mathcal{F}), \forall c \in C : \mathbb{E}(|c||\mathcal{F}) \leq f$, then for any $\epsilon \in L^0_+(\mathcal{F})$ taking $a = f/\epsilon$ follows $\mathbb{P}(|c| \geq a|\mathcal{F}) \leq \epsilon$). If in Lemma 4.5.2 the sequence $\{\xi_n\}_n$ was more generally real valued then the existence of a sequence in the convex hull of its tails, convergent almost surely to a finite limit follows from $conv\{|\xi_n| : n \in \mathbb{N}\}$ (or equivalently both $conv\{\xi_n^+ : n \in \mathbb{N}\}$ and $conv\{\xi_n^- : n \in \mathbb{N}\}$) being $\mathcal{F}$-conditionally bounded in probability. For vector valued sequences we may use these criteria component-wise.

4.6 General attainability results

Let us return to our Principal-Agent problem. By the recursions already derived we have to deal with (single-period) conditional optimization problems, both for the Principal and the Agent, for which we just gave a series of results, tools and nomenclature at our disposal.

Remark 4.6.1 *Because we will be dealing with (conditional) optimization problems without constraints, the idea will be to show that the objective functions to be maximized have super-level sets (e.g. of the form $\{objective.function \geq 0\}$) bounded in appropriate senses, among other properties. We should mention that there are other ways one may approach the existence of an optimizer for a convex optimization problem, for instance by showing that it is the (parametric) dual of a problem for which suitable regularity (i.e. qualification conditions) hold, or by Hahn-Banach separation results with which one may prove the non-emptiness of the topological super-differential of the concave conjugate of the objective function at the origin.*

We start with Agent's problem. First let us call

$$G(t, X, \beta) := \operatorname*{ess\,sup}_{A_t \in L^0(\mathcal{F}_t)^N} \left\{ -c_t(A_t)\Delta t + U_t^a \left(X + \beta A_t \Delta \tilde{P}_{t+1} \right) \right\}.$$

This function equals the right hand-side in Agent's recursion (4.4.2), if we take $(X, \beta) = (H_{t+1}, \beta_t)$.

We will first give a general attainability result for G, as it will spare us some repetitive work later. For simplicity, let us call:

$$g_t(A) = -c_t(A)\Delta t + U_t^a(X + \beta A \Delta \tilde{P}).$$

This function has to be maximized over all $A \in [L^0(\mathcal{F}_t)]^N$. Under the usual assumptions g is automatically an $\mathcal{F}_t-$concave function, and hence stable (see Definition 4.5.2). The main difficulty for the conditional optimization problem of determining G is that we need to reduce the problem to a bounded (concretely, $L^0(\mathcal{F}_t)-$bounded) set. The following concept will prove useful for that matter:

Definition 4.6.1 *A family of time-consistent utility functionals $(U_t)_t$ is said to be Sensitive to Large Losses if $\lim_{\lambda \to \infty} U_0(\lambda x) = -\infty$ for every x such that $\mathbb{P}(x < 0) > 0$.*

The idea behind this definition is that dissatisfaction from losses outpace asymptotically the satisfaction from gains. This property is fulfilled by entropic families of functionals as well as by most optimized certainty equivalents pasted together (see Example 4.3.1); we refer to Cheridito et al. [a] and Cherny and Kupper [2007] for more on this.

The following result deals with G, and its proof is partly inspired by [Cheridito et al., a, Lemma B.1]:

Theorem 4.6.1 *Let U^a satisfy the usual assumptions. For a fixed t assume that U_t^a is L^0-upper-semicontinuous and $X \in dom(U_t^a)$. Then, under any of the next three conditions the problem $G(t, X, \beta) = \operatorname{ess\,sup}_{A\in[L^0(\mathcal{F}_t)]^N} g_t(A)$ is attained:*

- *U^a is Sensitive to Large Losses, $c_t(\cdot) \geq K' + \lambda|\cdot|$ with $\lambda > 0$ and $X \in dom(U_0^a)$,*
- *$U_t^a(\cdot) \leq K + \mathbb{E}[\cdot|\mathcal{F}_t]$, $E(X_+|\mathcal{F}_t) < \infty$ and $c(\cdot) \geq K' + \lambda|\cdot|$ with $\lambda\Delta t > \left|\beta\mathbb{E}[\Delta\tilde{P}|\mathcal{F}_t]\right|$ a.s.,*
- *$U_t^a(\cdot) \leq K + \mathbb{E}[\cdot|\mathcal{F}_t]$, $E(X_+|\mathcal{F}_t) < \infty$, $\Delta P_{t+1} \in L^1_{\mathcal{F}_t}(\mathcal{F}_{t+1})$ and $\lim_{|a|\to\infty} \frac{c(a)}{|a|} = +\infty$,*

for generic constants $K' \in \mathbb{R}$ and $K \in L^0(\mathcal{F}_t)$.

Proof. We intend to apply Theorem 4.5.3. Let us call $\psi(A) = \beta A diag(P)^{-1}$. Then $g_t(A) = U_t^a(X + \psi(A)\Delta P) - c(A)\Delta t$. Evidently

$$\operatorname*{ess\,sup}_{A\in[L^0(\mathcal{F})]^N} g_t(A) = \operatorname*{ess\,sup}_{A\in\Lambda} g_t(A),$$

where $\Lambda = \{A : g_t(A) \geq g_t(0)\}$. Clearly the set Λ is sequentially closed, L^0-convex, contains the origin and is $\sigma-$stable. Now we define

$$M = \{A \in [L^0(\mathcal{F})]^N : \psi(A)\Delta P_{t+1} = 0\},$$

and $M^\perp$ its orthogonal (almost pointwise) complement. We are ready to prove that Λ is L^0-bounded under each of the conditions in the statement of the present Theorem.

For the first condition:

Take initially $0 \neq A \in M^\perp$, hence $\psi(A)\Delta P \neq 0$. Because there is No-Arbitrage in the market we have that $\mathbb{P}(\psi(A)\Delta P < 0) > 0$, from which there exists a $\delta > 0$ such that $\psi(A)\Delta P < -\delta$ on a certain $\Omega_1 \subset \mathcal{F}_{t+1}$ with positive measure. On the other hand, since X is a.s. finite we can always find a $C \in \mathbb{R}$ such that the set $\{X \leq C\}$ has measure arbitrarily close to 1. From this, if C is large enough the set $\Omega_2 := \{X \leq C\} \cap \Omega_1 \subset \mathcal{F}_{t+1}$ has positive measure. This implies the existence of an $m \in \mathbb{N}$ such that

$$\mathbb{P}(X^+ + m\psi(A)\Delta P < 0) > 0.$$

Indeed, otherwise a.s. for every $q \in \mathbb{N}$: $X^+ \geq q(-\psi(A)\Delta P)$, from which a.s. on Ω_2 and for all q we would have $C \geq q\delta$, which is a contradiction. The random variable $Y := (X^+ + m\psi(A)\Delta P)/m$ satisfies $\mathbb{P}(Y < 0) > 0$ and hence $U_0^a(nY) \to -\infty$ as n goes to infinity by sensitivity to large losses. On the other hand:

$$\forall n \geq m : X + n\psi(A)\Delta P \leq nY \text{ and thus } U_0^a(X + n\psi(A)\Delta P) \to -\infty.$$

This shows that for any given element $a + A \in \Lambda$, with $a \in M, A \in M^\perp$, there is a large n such that $n(A + a) \notin \Lambda$ as long as $A \neq 0$. Indeed, if the contrary were the case, by bounding $-c \leq -K'$, definition of Λ and recalling time-consistency (and monotonicity) it would follow that

$$-K'\Delta t + U_0^a(X + n\psi(A)\Delta P) \geq U_0^a(X) - c(0)\Delta t,$$

whose left-hand side tends to $-\infty$ by the previous paragraph, yielding a contradiction.

For the case $A = 0$ we easily see that $g_t(na) = U_t(X) - c(na)\Delta t$ tends to $-\infty$ on a non-negligible set (by the growth condition of c) as soon as $a \neq 0$, proving that $0 \neq a \in \Lambda \cap M$ implies $na \notin \Lambda$ for large n. By Lemma 4.5.2 we get that Λ is L^0-bounded.

As for the second condition:

We notice that for $(a, A) \in M \times M^\perp$:

$$\begin{aligned} g_t(nA + na) &= U_t^a(X + \psi(nA)\Delta P) - c(nA + na)\Delta t \\ &\leq K + \mathbb{E}[X_+|\mathcal{F}_t] + n\mathbb{E}[\psi(A)\Delta P|\mathcal{F}_t] - c(nA + na)\Delta t. \end{aligned} \tag{4.6.1}$$

Using that $A \in L^0(\mathcal{F}_t)^N$ and Cauchy-Schwarz applied pointwise, the sum of the last two terms is bounded from above by

$$n|A||\beta\mathbb{E}[\Delta\tilde{P}|\mathcal{F}_t]| - n|A|\lambda\Delta t - n|a|\lambda\Delta t - K'\Delta t.$$

Hence we see that $g_t(nA + na)$ tends to $-\infty$ on a non-negligible as n grows, under the assumption made on λ. But $g_t(0) > -\infty$ and again the set Λ must be L^0-bounded by Lemma 4.5.2.

As for the third condition, we see first from equation (4.6.1) that

$$g_t(nA + na) \leq K + \mathbb{E}[X_+|\mathcal{F}_t] + n|A||\beta||\mathbb{E}[\Delta\tilde{P}|\mathcal{F}_t]| - c(nA + na)\Delta t.$$

If $A = 0$ then $a \neq 0$ and we get that $g_t(na) \to -\infty$ on a non-negligible. If now $A = 0$ does not hold, we have that in the set where A is not null:

$$n|A||\beta||\mathbb{E}[\Delta\tilde{P}|\mathcal{F}_t]| - c(nA + na)\Delta t \leq n|A| \left[|\beta||\mathbb{E}[\Delta\tilde{P}|\mathcal{F}_t]| - \frac{c(nA + na)}{n|A|} \right].$$

Since $|\mathbb{E}[\Delta\tilde{P}|\mathcal{F}_t]|$ is a.s. finite we see that the majorizing term tends to $-\infty$ on a non-negligible set and thus $g(nA+na)$ does it likewise. As for the previous case, this implies in the end that the set Λ must be L^0-bounded.

Hence under any of the three conditions Theorem 4.5.3 applies for $\operatorname{ess\,sup}_{A\in\Lambda} g(A)$, yielding attainability. ■

Remark 4.6.2 *Regarding the second and third conditions in the previous result: if a representation as $U_t^a(X) = \operatorname{ess\,inf}_{Z\in\mathcal{W}}\{\mathbb{E}[ZX|\mathcal{F}_t] + \alpha(Z)\}$ was available (see section 4.5), then we could have obtained the requirement $U^a(\cdot) \leq K + \mathbb{E}[\cdot|\mathcal{F}_t]$ under the condition $1 \in dom(\alpha)$. This is the case if U arises as an optimized certainty equivalent (see example 4.3.2) for which $U^a(\cdot) = \operatorname{ess\,sup}_s\{s - \mathbb{E}[H(s - \cdot)|\mathcal{F}]\}$, because then the corresponding α in its variational representation coincides on $\mathcal{W}$ (as can be found in Cheridito et al. [a]) with $\mathbb{E}[H^*(\cdot)|\mathcal{F}]$, and $H^*(1) = 0$ by assumption on H (as usual*

H^* stands for the Fenchel conjugate of H). A slight generalization of this argument would have been to bound $U_t^a(\cdot) \leq K + \mathbb{E}[\cdot Z|\mathcal{F}_t]$ for some $Z \in \mathcal{W} \cap dom(\alpha)$, but then the condition $E(ZX_+|\mathcal{F}_t) < \infty$ has to be generally taken into account, the criterion relating λ and ΔP has to be adjusted to $\lambda \Delta t > \left|\beta \mathbb{E}[Z\Delta\tilde{P}|\mathcal{F}_t]\right|$ a.s. in the second case, and in the third case $Z\Delta P \in L^1_{\mathcal{F}_t}(\mathcal{F}_{t+1})$ should hold. Such an argument will be made in Theorem 4.6.2.

Thus, for a fixed contract and time, Agent's one-step problems are attainable under the conditions of the following direct corollary:

Corollary 4.6.1 *Consider the one-step conditional optimization problem of the Agent, as in* (4.4.2)*. Let U_t^a satisfy the usual assumptions and further be L^0-upper semicontinuous, and $H_{t+1} \in dom(U_t^a)$. Then, under any of the next three conditions Agent's problem* (4.4.2) *at time t is attained:*

- *U^a is Sensitive to Large Losses, $c_t(\cdot) \geq K' + \lambda|\cdot|$ with $\lambda > 0$ and $H_{t+1} \in dom(U_0^a)$,*
- *$U_t^a(\cdot) \leq K + \mathbb{E}[\cdot|\mathcal{F}_t]$, $E([H_{t+1}]_+|\mathcal{F}_t) < \infty$ and $c_t(\cdot) \geq K' + \lambda|\cdot|$ with $\lambda > \frac{|\beta|}{\Delta t}|\mathbb{E}[\Delta\tilde{P}_{t+1}|\mathcal{F}_t]|$ a.s.,*
- *$U_t^a(\cdot) \leq K + \mathbb{E}[\cdot|\mathcal{F}_t]$, $E([H_{t+1}]_+|\mathcal{F}_t) < \infty$, $\Delta P_{t+1} \in L^1_{\mathcal{F}_t}(\mathcal{F}_{t+1})$ and $\lim\limits_{|a|\to\infty} \frac{c_t(a)}{|a|} = +\infty$,*

for generic constants $K' \in \mathbb{R}$ and $K \in L^0(\mathcal{F}_t)$.

We move on to the attainability of Principal's unconstrained one-step problems, as in Theorem 4.4.2. Let us call:

$$V_t(A, \Gamma) := U_t^a(\Gamma) - c_t(A)\Delta t + U_t^p\left(h_{t+1} + A\Delta\tilde{P}_{t+1} - \Gamma\right),$$

and remember from Remark 4.4.2 that $V_t(A,\Gamma) \in L^0(\mathcal{F}_t)$ and from Remark 4.4.4 that we may consider, and will do it for the rest of the chapter, $U_t^p : L^0(\mathcal{F}_T) \mapsto L^0(\mathcal{F}_t)$ satisfying the usual assumptions w.r.t. $\mathcal{F}$.

Let us introduce the following assumption, which will hold for the rest of this work.

Assumption 18 *$dom(U_t^p) \cup dom(U_t^a) \subset \{Y \in L^0(\mathcal{F}_T) : \mathbb{E}[Y_-|\mathcal{F}_t] < \infty$, a.s.$\}$ and $dom(U_t^p) \cap dom(U_t^a) \neq \emptyset$.*

With this assumption, if $\mathbb{E}(|\Gamma||\mathcal{F}_t)$ is not a.s. finite, then the random variables of the form $U_t^a(\Gamma) + U_t^p(E - \Gamma)$ will take the value $-\infty$. Therefore when optimizing V over Γ it will be assumed that $\mathbb{E}(\Gamma_\pm|\mathcal{F}_t)$ are finite (i.e. $\Gamma \in L^1_{\mathcal{F}_t}(\mathcal{F}_{t+1})$). This justifies that w.l.o.g. we can further assume $\mathbb{E}[\Gamma|\mathcal{F}_t] = 0$ when maximizing V, by translation invariance. This assumption holds automatically for entropic functionals, for instance.

We provide now a general attainability result for Principal's unconstrained one-step problems. This improves on Theorem 4.4.2 in that, under further technical conditions, the attainability and finiteness hypotheses therein are now guaranteed.

Theorem 4.6.2 *Suppose that beyond satisfying the usual properties and Assumption 18, U_t^a and U_t^p are L^0-upper-semicontinuous, satisfy $dom(U_t^p), dom(U_t^a) \subset L^1_{\mathcal{F}_t}(\mathcal{F}_T)$ and have the following representations:*

$$\begin{aligned} U_t^p(X) &= \operatorname*{ess\,inf}_{Z\in\mathcal{W}} \{\mathbb{E}[ZX|\mathcal{F}_t] + \alpha_t^p(Z)\} \ \textit{for}\ X \in L^1_{\mathcal{F}_t}(\mathcal{F}_{t+1}), \\ U_t^a(X) &= \operatorname*{ess\,inf}_{Z\in\mathcal{W}} \{\mathbb{E}[ZX|\mathcal{F}_t] + \alpha_t^a(Z)\} \ \textit{for}\ X \in L^1_{\mathcal{F}_t}(\mathcal{F}_{t+1}), \end{aligned} \tag{4.6.2}$$

where $\alpha_t^a, \alpha^p : L^\infty_{\mathcal{F}_t}(\mathcal{F}_{t+1}) \to \overline{L}(\mathcal{F}_t)$ and $\mathcal{W} := \{Z \in L^\infty_{\mathcal{F}_t}(\mathcal{F}_{t+1}) : Z \geq 0, \mathbb{E}[Z|\mathcal{F}_t] = 1\}$.
Assume that

$$K_t^p := \operatorname*{ess\,sup}_{Z\in\mathcal{W}\cap[0,2]} \alpha^p(Z) \in L^0(\mathcal{F}_t) \ \textit{and}\ K_t^a := \operatorname*{ess\,sup}_{Z\in\mathcal{W}\cap[0,2]} \alpha^a(Z) \in L^0(\mathcal{F}_t).$$

Finally suppose that $h_{t+1} \in dom(U_t^p), \Delta P_{t+1} \in L^1_{\mathcal{F}_t}(\mathcal{F}_T)$, and c_t satisfy either:

- *$c_t(\cdot) \geq K + \lambda|\cdot|$ and a.s. $\lambda\Delta t > \left|\tilde{Z}\mathbb{E}[\Delta\tilde{P}_{t+1}|\mathcal{F}_t]\right|$, some $\tilde{Z} \in dom(\alpha^p) \cap dom(\alpha^a) \cap \mathcal{W} \cap L^0(\mathcal{F}_t)$*
- *$\lim\limits_{|a|\to\infty} \frac{c_t(a)}{|a|} = +\infty$.*

Then $\Sigma := \operatorname*{ess\,sup}_{(A,\Gamma)\in[L^0(\mathcal{F}_t)]^N\times L^1_{\mathcal{F}_t}(\mathcal{F}_{t+1})} V_t(A,\Gamma) \in L^0(\mathcal{F}_t)$ and this problem has an optimal solution and thus at time t we have that $\beta_t = 1$ is optimal.

Proof. Define $\mathcal{S} = \{(A,\Gamma) \in L^0(\mathcal{F}_t)^N \times Q : V(A,\Gamma) \geq V(0,0)\}$, where

$$Q := \left\{\Gamma \in L^1_{\mathcal{F}_t}(\mathcal{F}_{t+1}) : \mathbb{E}[\Gamma|\mathcal{F}_t] = 0\right\}.$$

Notice that in computing Σ we may assume that (A,Γ) belong to $\mathcal{S}$, since $\mathcal{F}_t$-measurable components of Γ cancel out in V. In a first step, we will show that the set

$$\mathcal{S}^A := \{A \in L^0(\mathcal{F}_t)^N : \exists\Gamma \in Q \text{ such that } (A,\Gamma) \in \mathcal{S}\},$$

of A's in $\mathcal{S}$, is $L^0(\mathcal{F}_t)-$bounded. First, notice $V(0,0) = -c(0)\Delta t + U_t^p(h_{t+1}) \in L^0(\mathcal{F}_t)$. Taking $\tilde{Z} \in dom(\alpha^p) \cap dom(\alpha^a) \cap \mathcal{W} \cap L^0(\mathcal{F}_t)$ (by assumption, 1 belongs to it, for instance), then $U_t^a(\Gamma) \leq \alpha^a(\tilde{Z}) + \tilde{Z}\mathbb{E}[\Gamma|\mathcal{F}_t]$ and

$$U_t^p(h + A\Delta\tilde{P} - \Gamma) \leq \alpha^p(\tilde{Z}) + \tilde{Z}\mathbb{E}[h|\mathcal{F}_t] - \tilde{Z}\mathbb{E}[\Gamma|\mathcal{F}_t] + \tilde{Z}\mathbb{E}[A\Delta\tilde{P}_{t+1}|\mathcal{F}_t].$$

For $\Gamma \in Q$ the term $\mathbb{E}[\Gamma|\mathcal{F}_t]$ vanishes and hence

$$V(0,0) \leq \alpha^p(\tilde{Z}) + \alpha^a(\tilde{Z}) + \tilde{Z}\mathbb{E}[h|\mathcal{F}_t] + |A||\tilde{Z}\mathbb{E}[\Delta\tilde{P}_{t+1}|\mathcal{F}_t]| - c(A)\Delta t.$$

Since we clearly have that $\mathcal{S}^A$ is σ-stable, we can use Lemma 4.5.1 to prove what we want. Indeed, if $\mathcal{S}^A$ were not $L^0(\mathcal{F}_t)-$bounded, we would get the existence of a non-negligible set $\tilde{\Omega}$ and a sequence $\{A_n\} \subset \mathcal{S}^A$ such that $|A_n| \geq n$ on $\tilde{\Omega}$. It is a simple exercise, in the spirit of the proof of Theorem 4.6.1, to check that under either growth condition stated for c (one would additionally require a.s. $\lambda\Delta t > \left|\tilde{Z}\mathbb{E}[\Delta\tilde{P}_{t+1}|\mathcal{F}_t]\right|$ for the first one), we would have that $V(0,0) = -\infty$ on a non-negligible set, which is a contradiction to our hypotheses. Thus $\mathcal{S}^A$ must be $L^0(\mathcal{F}_t)-$bounded.

Now, call $\mathcal{S}^\Gamma$ the set of Γ's in $\mathcal{S}$ (this is its projection in the "Γ" coordinate); we will prove it to be bounded in $L^1_{\mathcal{F}_t}(\mathcal{F}_T)$. Fix such a Γ and define

$$Z^a := \mathbb{1}_{\Gamma\leq 0} + \mathbb{P}(\Gamma > 0|\mathcal{F}_t) \text{ and } Z^p := \mathbb{1}_{\Gamma>0} + \mathbb{P}(\Gamma \leq 0|\mathcal{F}_t).$$

Notice that $Z^a, Z^p \in L^\infty(\mathcal{F}_T) \cap [0,2]$. Since $\Gamma \in Q$ we see that

$$\mathbb{E}[Z^a\Gamma|\mathcal{F}_t] = -\mathbb{E}[(\Gamma)_-|\mathcal{F}_t] \text{ and } \mathbb{E}[Z^p\Gamma|\mathcal{F}_t] = \mathbb{E}[(\Gamma)_+|\mathcal{F}_t].$$

Clearly also $\mathbb{E}[Z^a|\mathcal{F}_t] = \mathbb{E}[Z^p|\mathcal{F}_t] = 1$, implying that $Z^{a,p} \in \mathcal{W}$ and thus $\alpha^p(Z^p) \leq K^p$ and $\alpha^a(Z^a) \leq K^a$. We hence obtain that $U^a(\Gamma) \leq -\mathbb{E}[(\Gamma)_-|\mathcal{F}_t] + K^a$ and

$$\begin{aligned} U^p(h + A\Delta\tilde{P} - \Gamma) &\leq \mathbb{E}[Z^p(h + A\Delta\tilde{P})|\mathcal{F}_t] - \mathbb{E}[\Gamma_+|\mathcal{F}_t] + K^p \\ &\leq 2\mathbb{E}[|h||\mathcal{F}_t] + 2|A|\mathbb{E}[|\Delta\tilde{P}||\mathcal{F}_t] - \mathbb{E}[\Gamma_+|\mathcal{F}_t] + K^p \\ &\leq N - \mathbb{E}[\Gamma_+|\mathcal{F}_t], \end{aligned}$$

with $N \in L^0(\mathcal{F}_t)$ by assumption and the fact that the A's have been proven to be $L^0(\mathcal{F}_t)-$bounded. Therefore for $(A,\Gamma) \in \mathcal{S}$ one has

$$V(0,0) \leq V(A,\Gamma) \leq N + K^a - E[(\Gamma)_-|\mathcal{F}_t] - \mathbb{E}[\Gamma_+|\mathcal{F}_t] - c_t(A)\Delta t \leq \tilde{K} - E[|\Gamma||\mathcal{F}_t],$$

where $\tilde{K} \in L^0(\mathcal{F}_t)$ and upon using again the boundedness of $\mathcal{S}^A$ (and the continuity of $c_t(\cdot)$). This means of course that $\mathcal{S}^\Gamma$ is bounded in $L^1_{\mathcal{F}_t}(\mathcal{F}_T)$.

Recall that $\Sigma := \operatorname{ess\,sup}_{(A,\Gamma)\in\mathcal{S}} V(A,\Gamma)$ and notice that $\mathcal{S}$ is directed upwards. Indeed if $V(A^1,\Gamma^1) \geq V(0,0)$ and $V(A^2,\Gamma^2) \geq V(0,0)$ define $\xi = \{V(A^1,\Gamma^1) \geq V(A^2,\Gamma^2)\}$ and $(A,\Gamma) = (A^1,\Gamma^1)\mathbb{1}_\xi + (A^2,\Gamma^2)\mathbb{1}_{\xi^c}$. Then

$$V(A,\Gamma) = \max\{V(A^1,\Gamma^1), V(A^2,\Gamma^2)\} \geq V(0,0),$$

thanks to the terms in V being $\mathcal{F}_t-$stable and $\xi \in \mathcal{F}_t$. Therefore we have $\Sigma = \operatorname{ess\,sup}_{A,\Gamma} V(A,\Gamma) = lim_n V(A^n,\Gamma^n)$ and the limit is increasing. Take then $(A_n,\Gamma_n) \in \mathcal{S}$ such that $V(A_n,\Gamma_n)$ increases to Σ. Applying Lemma 4.5.2 iteratively (component-wise for the A_n's and then to the Γ_n's, see discussion around the mentioned lemma), we find the existence of non-negative real numbers $\{\lambda_i^n\}$ with $\sum_{i\geq n}\lambda_i^n = 1$, and random variables $\Gamma^* \in L^0(\mathcal{F}_{t+1})$ and $A^* \in L^0(\mathcal{F}_t)^N$ such that $\tilde{\Gamma}_n = \sum_{i\geq n}\lambda_i^n\Gamma_i \to \Gamma^*$ and $\tilde{A}_n = \sum_{i\geq n}\lambda_i^n A_i \to A^*$ almost surely. Notice that $(\tilde{A}_n,\tilde{\Gamma}_n) \in \mathcal{S}$ by convexity. What is more

$$\Sigma = \lim V(A_n,\Gamma_n) = \lim_n \sum_{i\geq n}\lambda_i^n V(A_i,\Gamma_i) \leq \limsup_n V(\tilde{A}_n,\tilde{\Gamma}_n),$$

where for the second equality it was used (a.s.) that real convergent sequences remain converging under convex combinations of its tails, and for the inequality that V is concave.

By using (sequential) L^0-uppersemicontinuity of each of the terms in V, we get that $\Sigma \leq V(A^*, \Gamma^*)$ and hence we have equality and therefore attainment. This shows that $\Sigma < \infty$ by properness of the U's. Finally by Theorem 4.4.2 we conclude that $\beta = 1$ is optimal and Principal's one-step problem is attained. ■

Remark 4.6.3 *From Theorem 4.5.4, we see that the variational representations* (4.6.2) *of U_t^p and U_t^a are consequence of them being proper functions from $L^1_{\mathcal{F}_t}(\mathcal{F}_{t+1})$ to $\underline{L}(\mathcal{F}_t)$, monotone, $\mathcal{F}_t-$translation invariant and $L^1_{\mathcal{F}_t}(\mathcal{F}_{t+1})-$upper semicontinuous. This last property is implied by their L^0-upper semicontinuity, as is proven in Proposition 4.5.1. We remark that if U_t^p has initially a variational representation with respect to $\mathcal{F}_t^A$, then $U_t^p(L^0(\mathcal{F}_T)) \subset \underline{L}^0(\mathcal{F}_t)$ is trivially satisfied. On the other hand, under the conditions for $K^{a,p}$ we have automatically that $dom(\alpha^p) \cap dom(\alpha^a) \cap \mathcal{W} \cap L^0(\mathcal{F}_t)$ is not empty since* 1 *would belong to it.*

Remark 4.6.4 *Building on Remark 4.6.2, the conditions in Theorem 4.6.2 concerning the $K^{a,p}$ are automatically satisfied if the utility functionals are of the form of optimized certainty equivalents for which the corresponding H's satisfy $H^*(0), H^*(2) \in \mathbb{R}$. In Theorem 4.6.5 we shall find a simple instance where also the other finiteness/integrability conditions are satisfied for these type of functionals.*

Remark 4.6.5 *For sake of generality we chose to work in the biggest conditional (loc. convex) space of L^p-type, this is, the conditional L^1 space. Had we worked with smaller subspaces, we would have had more tools at hand to prove the attainability of Principal's one-step problems. For instance, a natural space associated to an optimized certainty equivalent should be a conditional Orlicz space (related to the corresponding H function). We chose however not to limit the scope of utility functionals for which the theory would be applicable to, and thus work with conditional L^1.*

It is well known that if the utility functionals are of a similar structure, i.e. when they are a re-scaling of each other, a more explicit treatment of equilibrium/risk-sharing problems becomes available (see e.g. Barrieu and Karoui [2005] or the seminal work Borch [1960]). Along these lines we prove the following result:

Theorem 4.6.3 *Suppose there is a family of utility functionals U_t, satisfying the usual assumptions and being L^0-upper-semicontinuous, such that $U_t^i(\cdot) = \frac{1}{\gamma^i} U_t(\gamma^i \cdot)$ for $i = a, p$ (where the γ^is are positive). Let us define $\hat{\gamma} = \frac{\gamma^a\gamma^p}{\gamma^a+\gamma^p} = \frac{1}{1/\gamma^a+1/\gamma^p}$. Assuming that for a fixed time t any of the following holds:*

- *the family U is Sensitive to Large Losses, $c_t(\cdot) \geq K' + \lambda|\cdot|$ with $\lambda > 0$, and $\hat{\gamma}h_{t+1} \in dom(U_0) \cap dom(U_t)$*
- *$U_t(\cdot) \leq K + \mathbb{E}[\cdot|\mathcal{F}_t]$, $E([h_{t+1}]_+|\mathcal{F}_t) < \infty$, $c_t(\cdot) \geq K' + \lambda|\cdot|$ with $\lambda\Delta t > \left|\mathbb{E}[\Delta\tilde{P}_{t+1}|\mathcal{F}_t]\right|$, and $\hat{\gamma}h_{t+1} \in dom(U_t)$*
- *$U_t(\cdot) \leq K + \mathbb{E}[\cdot|\mathcal{F}_t]$, $E([h_{t+1}]_+|\mathcal{F}_t) < \infty$, $\Delta P_{t+1} \in L^1_{\mathcal{F}_t}(\mathcal{F}_{t+1})$, $\lim\limits_{|a|\to\infty} \frac{c_t(a)}{|a|} = +\infty$, and $\hat{\gamma}h_{t+1} \in dom(U_t)$,*

then Principal's one-step problem (at time t) has as solution:

$$\beta = 1 \text{ and } \Gamma^* = \frac{\gamma^p}{\gamma^a + \gamma^p}(h_{t+1} + A^*\Delta\tilde{P}_{t+1}),$$

for any optimal action A^* of the Agent, which attains:

$$\operatorname*{ess\,sup}_A \left\{ -c_t(A)\Delta t + \frac{1}{\hat{\gamma}} U_t\left(\hat{\gamma}[h_{t+1} + A\Delta\tilde{P}_{t+1}]\right) \right\}.$$

Proof. By Theorem 4.4.2, Principal's problem reduces to maximizing $-c_t(A)\Delta t + U_t^a(\Gamma) + U_t^p(h_{t+1} + A^*\Delta\tilde{P} - \Gamma)$ over $A \in L^0(\mathcal{F}_t)^N$ and $\Gamma \in L^0(\mathcal{F}_{t+1})$. Consider first A fixed, and call $x := h + A\Delta\tilde{P}_{t+1}$. Then by concavity:

$$\frac{1}{\gamma^a}U_t(\gamma^a\Gamma) + \frac{1}{\gamma^p}U_t(\gamma^p[x-\Gamma]) = \frac{1}{\hat{\gamma}}\left\{ \frac{\hat{\gamma}}{\gamma^a}U_t(\gamma^a\Gamma) + \frac{\hat{\gamma}}{\gamma^p}U_t(\gamma^p[x-\Gamma]) \right\} \leq \frac{1}{\hat{\gamma}}U_t(\hat{\gamma}x).$$

Hence $\operatorname{ess\,sup}_\Gamma \frac{1}{\gamma^a}U_t(\gamma^a\Gamma) + \frac{1}{\gamma^p}U_t(\gamma^p[x-\Gamma]) \leq \frac{1}{\hat{\gamma}}U_t(\hat{\gamma}x)$. On the other hand, taking $\Gamma^* = \frac{\gamma^p}{\gamma^a+\gamma^p}x$ it follows that $\frac{1}{\gamma^a}U_t(\gamma^a\Gamma^*) + \frac{1}{\gamma^p}U_t(\gamma^p[x-\Gamma^*]) = \frac{1}{\hat{\gamma}}U_t(\hat{\gamma}x)$. Therefore this Γ^* attains the essential supremum above. Thus, one can replace this again in Principal's problem w.r.t. variable A to get that she must maximize

$$w(A) := \left\{ -c_t(A)\Delta t + \frac{1}{\hat{\gamma}} U_t\left(\hat{\gamma}[h_{t+1} + A\Delta\tilde{P}_{t+1}]\right) \right\}. \tag{4.6.3}$$

If this problem is attained at a certain A^*, the previous argument shows that $\Gamma^* = \frac{\gamma^p}{\gamma^a+\gamma^p}(h_{t+1} + A^*\Delta\tilde{P}_{t+1})$ would be optimal.

The problem (4.6.3) is of the same form as that analyzed in Theorem 4.6.1, simply replacing U^a by $\frac{1}{\hat{\gamma}}U_t(\hat{\gamma}\cdot)$, calling $X = h$ and taking $\beta = 1$. So applying this result, under the corresponding conditions, we get the existence of A^* such that $w(A^*) = \operatorname{ess\,sup}_A w(A)$. Finally, because the one-step unconstrained problem is attained, Theorem 4.4.2 shows that taking $\beta = 1$, A^* and Γ^* at time t yields an optimal one-step decision. ∎

The following Lemma gives another sufficient condition for the attainability of Principal's problem, providing an alternative to the more technical Theorem 4.6.2. The point here is to assume a certain boundedness from the outset; thus no growth condition has to be required in principle:

Lemma 4.6.1 *Suppose that the usual assumptions on U^a and U^p and their L^0-upper-semicontinuity. A sufficient condition for $h_t = \Sigma$ is the existence of $r \in L^0(\mathcal{F}_t)$ and $\tilde{r} \in L^0(\mathcal{F}_{t+1})$, both a.s. positive, such that for every $(A,\Gamma) \in L^0(F_t)^N \times Q$ satisfying $V(A,\Gamma) \geq V(0,0)$ necessarily $|A| \leq r$ and $|\Gamma| \leq \tilde{r}$, where $Q = \{\Gamma : \mathbb{E}[\Gamma|\mathcal{F}_t] = 0\}$. If this is the case, then:*

$$h_t = V_t(A^*, \Gamma^*) \text{ for some } (A^*, \Gamma^*) \in \mathcal{C}(1).$$

Proof. In light of Theorem 4.4.2 we only need to check that the unconstrained problem $\operatorname{ess\,sup}_{A,\Gamma} V_t(A,\Gamma)$ is attained. Let us call

$$\mathcal{S} := \{(A,\Gamma) \in L^0(F_t)^N \times L^0(F_{t+1}) : V_t(A,\Gamma) \geq V_t(0,0)\}.$$

This set is oriented upwards, by the same argument as in the proof of Theorem 4.6.2. Therefore

$$\Sigma = \operatorname*{ess\,sup}_{A,\Gamma} V_t(A,\Gamma) = lim_n V(A^n,\Gamma^n),$$

and the limit is increasing. By translation invariance we may assume w.l.o.g. that $\Gamma_n \in Q$, as we have done before. From this, by assumption $\{\Gamma_n\}$ is bounded in $L^0(\mathcal{F}_{t+1})$ and therefore a Komlos argument (see e.g. Lemma 4.5.2 and surrounding discussion) gives the existence of scalars (this is important) λ^n_m such that $\tilde{\Gamma}_n := \sum_{m\geq n} \lambda^n_m \Gamma_m \in L^0(\mathcal{F}_{t+1})$ and so that $\tilde{\Gamma}_n \to \Gamma^*$ a.s. for a certain $\Gamma^* \in L^0(\mathcal{F}_{t+1})$. Call $\tilde{A}_n = \sum_{m\geq n} \lambda^n_m A_m$. By increasingness and concavity:

$$V(A_n,\Gamma_n) \leq \sum_{m\geq n} \lambda^n_m V(A_m,\Gamma_m) \leq V(\tilde{A}_n,\tilde{\Gamma}_n),$$

and so $\Sigma \leq \liminf_n V(\tilde{A}_n,\tilde{\Gamma}_n)$.

We have $\{\tilde{\Gamma}_n\} \in Q$, since $\Gamma_n \in Q$. The above inequalities also show that $V(0,0) \leq V(\tilde{A}_n,\tilde{\Gamma}_n)$, and thus by assumption $\{\tilde{A}_n\}$ is $L^0(\mathcal{F}_t)-$bounded, and hence from Theorem 4.5.1 (one could alternatively use a Komlos argument as well) we get the existence of $A^* \in L^0(\mathcal{F}_t)^N$ such that $\tilde{A}_{N_n} \to A^*$ a.s., for a randomized sequence $N_n \in \mathbb{N}(\mathcal{F}_t)$. It is clear that still $\tilde{\Gamma}_{N_n} \to \Gamma^*$. Notice that

$$V(\tilde{A}_{N_n},\tilde{\Gamma}_{N_n}) = \sum_{m\geq n} V(\tilde{A}_m,\tilde{\Gamma}_m)\mathbb{1}_{N_n=m} \geq \inf_{m\geq n} V(\tilde{A}_m,\tilde{\Gamma}_m),$$

and by taking limsup and then recalling the inequality derived for Σ, we see that $\Sigma \leq \limsup_n V(\tilde{A}_{N_n},\tilde{\Gamma}_{N_n}) \leq V(A^*,\Gamma^*)$, by upper-semicontinuity. Since $A^* \in L^0(\mathcal{F}_t)^N$ and $\Gamma^* \in L^0(\mathcal{F}_{t+1})$, this shows that Σ is attained and thus finite. ∎

Remark 4.6.6 *It is not difficult to find examples where the set of* Γ*s such that* $K \leq U^a(\Gamma) + U^p(-\Gamma)$*, is not "pointwise" (i.e.* $L^0(\mathcal{F}_{t+1})$*-) bounded, and where even the essential supremum and infimum of such family is plus and minus infinity respectively. This is why Lemma 4.6.1 cannot be always applied in this level of generality.*

Despite the previous comment, the latter lemma will prove very useful for the so-called case with *predictable representation property*, which we will introduce soon.

The most important results in this section were Theorems 4.6.2 and 4.6.3 for the attainability of Principal's one-step problems respectively without or with scalability of the utility functionals in a general filtration. Let us remark that in each case only the instantaneous (one-step) formulation of the (PA) problem was analyzed. So does that mean that the whole dynamic PA problem (in $\{0,\ldots,T\}$) was solved? In light of Theorem 4.4.1 the answer is positive, up to some technicalities. We know that what is needed is that at each time step the one-step problems be finite and attained. However, the output of the problem at step $t+1$ (that is H_{t+1} and h_{t+1}) enters into the problem at time t, and in order to apply the sufficient conditions we have, we need these random variables to be sufficiently integrable or belong to certain domains. To clarify this, the following summarizing result is in order (the usual hypotheses, L^0-upper semicontinuity and Assumption 18 also hold of course). We omit the obvious proof.

Theorem 4.6.4 *In the setting of generic dynamic utility functionals (see Theorem 4.6.2) satisfying for every t the assumptions on $K_t^{a,p}$ and such that the c_t's satisfy any of the two required assumptions therein for the whole time span, if $h_{t+1} \in dom(U_t^p)$ holds for every t, then the one-step problems have a solution and glueing them together yields a solution for the dynamic problem.*

In the setting of utility functionals of the same form, i.e. coming from scaling of a fixed one (see Theorem 4.6.3), and assuming the U_t's, c_t's and h_t's satisfy any of the three required assumptions therein for every time, then the one-step problems have a solution and glueing them together yields a solution for the dynamic problem.

In particular, in any of the previous cases, $\beta_t = 1$ for every t is optimal.

Finally, to illustrate how the assumptions in the previous theorem may be checked, we analyse the simpler case of certainty equivalent utility functionals and bounded prices:

Theorem 4.6.5 *In the case of certainty equivalent functionals of the same kind, in a market with bounded prices ($0 < p_- \leq P_t^i \leq p_+$ a.s.), and c super linear with a strong enough slope, the whole PA dynamic problem has a solution.*

Proof. By assumption U^p and U^a stem from $U_t(X) = \operatorname{ess\,sup}_s\{s - \mathbb{E}[H(s-X)|\mathcal{F}_t]\}$, for some normalized, convex and increasing function H. By previous remarks, this implies that $U_t(X) \leq \mathbb{E}[X|\mathcal{F}_t]$. It thus suffices to show that h_{t+1} is bounded by constants, in order to apply Theorem 4.6.4.

First of all $|\Delta\tilde{P}_{t+1}| = |diag(P_t)^{-1}\Delta P_{t+1}| \leq \frac{2p_+}{p_-}$. Now, since c_t is lower-semicontinuous, has super-linear growth and is convex, it has a global minimum (say at $a \in \mathbb{R}^N$). Clearly $(a, 0) \in \mathcal{C}_t(0)$, thus showing that

$$h_t \geq V_t(a,0) = -c(a)\Delta t + U_t^p(h_{t+1} + a \cdot diag(P_t)^{-1}\Delta P_{t+1}) \geq k + U_t^p(h_{t+1}),$$

for some constant k. Now, by backwards induction and since $h_T = 0$, we have that there is a constant α_- so that $h_{t+1}^p \geq \alpha_-$ for all t. On the other hand, by the upper bounds on U^p and U^a, we have that

$$\begin{aligned} V_t(A,\Gamma) &\leq -c(A)\Delta t + A\mathbb{E}[\Delta\tilde{P}|\mathcal{F}_t] + \mathbb{E}[h_{t+1}|\mathcal{F}_t] \\ &\leq -K - (\lambda - 2p_+/p_-)|A|\Delta t + \mathbb{E}[h_{t+1}|\mathcal{F}_t] \\ &\leq K'' + \mathbb{E}[h_{t+1}|\mathcal{F}_t], \end{aligned}$$

where $K'' \in \mathbb{R}$ and the super-linear growth of c (with λ large enough) was used. From here we get that $h_t \leq K'' + \mathbb{E}[h_{t+1}|\mathcal{F}_t]$ and again by backwards inductions follows that $h_{t+1}^p \leq \alpha_+$, for another constant. ∎

4.7 The case under the predictable representation property

Here we intend to apply Lemma 4.6.1 under the so-called Predictable Representation Property (PRP for short) assumption. The reason why the PRP-case works (without having to resort to Theorem 4.6.2) is that all of the one-step decision variables now will be measurable with respect to the same sigma-algebra, and hence the finite-dimensional conditional analysis ideas and results will apply more directly, providing for instance the "pointwise" boundedness of the decision variables in our one-step problems. We remark that in this section we aim to relax the condition involving the α's in Theorem 4.6.2 and at the same time provide less technical proofs without invoking explicitly the variational representation of the utility functionals.

In order to avoid repetitions, throughout this section we make the following assumption at every time: $c_t(\cdot) \geq K + \lambda|\cdot|$, the U_t's satisfy the usual properties, are L^0-upper-semicontinuous and Sensitive to Large Losses, and for the price increments we have $\mathbb{E}[|\Delta\tilde{P}_t||\mathcal{F}_t] < \infty$ a.s. Furthermore, we assume also throughout this section that the price process is a martingale (that is, that the reference measure is a martingale measure). Notice that because the output wealth of Agent's decisions do not enter directly into his utility function, the previous martingality assumption does not trivialize the problem at all.

We introduce now the defining assumption for the current analysis:

Assumption 19 *[Predictable Representation Property (PRP)]*
There is an $m < \infty$ and an $m-$dimensional martingale process M adapted to the original filtration $\mathcal{F}$, such that the $N + m =: D-$dimensional process $R := (P, M)$ has the following Predictable Representation Property, with unique decomposition, for every t:

$$L^0(\mathcal{F}_{t+1}) = \{z + Z^1 \cdot \Delta P_{t+1} + Z^2 \cdot \Delta M_{t+1} : z \in L^0(\mathcal{F}_t), Z^1 \in [L^0(\mathcal{F}_t)]^N, Z^2 \in [L^0(\mathcal{F}_t)]^m\}$$

Let us present an example of this notion:

Example 4.7.1 (Bernoulli Walk) *Consider in $\mathbb{R}^d$, d independent Bernoulli walks $w^1, \ldots, w^d$ on $\{0, h, 2h, \ldots, T\}$ starting at 0, such that $\mathbb{P}(\Delta w_t^i = \sqrt{h}) = \mathbb{P}(\Delta w_t^i = -\sqrt{h}) = \frac{1}{2}$. Assume that the filtration is the natural one associated to these processes. Clearly these increments are then centred, independent to each other and to their past, and have variance equal to h. However, they do not necessarily have the predictable representation property, unless $d = 1$. Yet, as shown in the appendix of Cheridito et al. [a] and calling $D = 2^d - 1$, there exists an adapted family $w^{d+1}, \cdots, w^D$ of likewise distributed random walks, such that the whole extended family $w^1, \ldots, w^D$ has increments uncorrelated to each other and independent from the past, and enjoys the Predictable Representation Property. We call this family in the following simply "the Bernoulli walk."*

Notice that under the PRP assumption $E(\cdot|\mathcal{F}_t)$ always exists. Moreover

$$\gamma\Delta R_{t+1} = 0 \text{ if and only if } \gamma = 0.$$

We can thus identify $\Gamma \in L^0(\mathcal{F}_{t+1})$ with $\gamma \in [L^0(\mathcal{F}_t)]^D$ by means of $\Gamma - \mathbb{E}[\Gamma|\mathcal{F}_t] = \gamma \cdot \Delta R$. We remark that in this context the PRP hypothesis is a type of market completion.

From now on, in this section, Assumption 19 is supposed to hold and we redefine $V_t, \mathcal{C}_t(\cdot)$ and the recursion for h in terms of $\gamma \in [L^0(\mathcal{F}_t)]^D$ by replacing Γ with $\gamma \cdot \Delta R$.

The next lemma shows that in the presence of PRP our utility functionals are affinely dominated by the conditional expectation thanks to sensitivity to large losses. Notice however that this does not make any of the sufficient conditions in Theorem 4.6.1 redundant, as a weaker condition on U^a requires stronger conditions on c.

Lemma 4.7.1 *Take $U_t : L^0(\mathcal{F}_T) \to \underline{L}(\mathcal{F}_t)$ (which stands for either U_t^a or U_t^p). Then there exists $K \in L^0(\mathcal{F}_t)$ such that:*

$$U_t(X) \leq K + \mathbb{E}[X|\mathcal{F}_t].$$

Proof. Let $\nabla = \{A \in L^0(\mathcal{F}_t)^D : U_t(A\Delta R_{t+1}) \geq 0\}$. The set ∇ is not-empty, sequentially closed, stable and conditionally convex. To prove that ∇ is $L^0(\mathcal{F}_t)-$bounded it suffices by Theorem 4.5.2 to prove that for every $A \in \nabla, A \neq 0$, there exists an n such that $nA \notin \Lambda$. Indeed, for such an A we have that $A\Delta R_{t+1}$ cannot be identically null, and since $\mathbb{E}[A\Delta R_{t+1}] = 0$, we have $A\Delta R_{t+1} < 0$ on a set with positive measure, from which by hypothesis $U_0(nA\Delta R_{t+1}) \to -\infty$ which implies $nA \notin \Gamma$ for large n, by time consistency. Therefore by Theorem (4.5.3) we have that $K := \operatorname{ess\,sup}_{A \in L^0(\mathcal{F}_t)^D} U_t(A\Delta R_{t+1}) \in L^0(\mathcal{F}_t)$. By the PRP then, $U_t(X) - \mathbb{E}[X|F_t] \leq K$. ■

To prove the attainability of Principal's problem under PRP the following Lemma will be useful:

Lemma 4.7.2 *Assume that $\mathbb{E}[h_{t+1}|\mathcal{F}_t] < \infty$. If $V_t(0,0) \in dom(U_0^p) \cup dom(U_0^a)$ then*

$$\left\{(A,\gamma) \in L^0(\mathcal{F}_t)^N \times L^0(\mathcal{F}_t)^D : V_t(A,\gamma) \geq V_t(0,0)\right\},$$

is $L^0(\mathcal{F}_t)$-bounded.

Proof. The set $\mathcal{S} := \left\{(A,\gamma) \in L^0(\mathcal{F}_t)^N \times L^0(\mathcal{F}_t)^D : V_t(A,\gamma) \geq V_t(0,0)\right\}$ is clearly $L^0(\mathcal{F}_t)-$convex and actually $\sigma-$stable thanks to remark 4.5.1. It is also sequentially closed, by sequential upper-semicontinuity of $V_t(\cdot,\cdot)$. As usual we call $\mathcal{S}^A$ and $\mathcal{S}^\gamma$ the projection of the set $\mathcal{S}$ onto its first and second coordinate. Notice that $\sigma-$stability of $\mathcal{S}^A$ is inherited from that of $\mathcal{S}$. In the following we aim at proving L^0-boundedness of $\mathcal{S}$.

We first prove that $\mathcal{S}^A$ is L^0-bounded. Using Lemma 4.7.1 we bound the U's by conditional expectations. Notice that the terms $\mathbb{E}(\gamma\Delta R|\mathcal{F}), \mathbb{E}(A\Delta\tilde{P}|\mathcal{F})$ vanish. Hence $V(A,\gamma) \leq -c(A)\Delta t$. Lemma 4.5.1 tells us that if $\mathcal{S}^A$ were not L^0-bounded there would exist a non-negligible set $\tilde{\Omega}$ and a sequence $\{A_n\} \subset \mathcal{S}^A$ such that $|A_n| \geq n$ on $\tilde{\Omega}$.

Therefore under the growth condition of c we get then that $V_t(0,0) = -\infty$ on $\tilde{\Omega}$, which is a contradiction. Therefore indeed $\mathcal{S}^A$ is L^0-bounded.

We will seek to apply Theorem 4.5.2 to get boundedness of $\mathcal{S}^\gamma$, but to this end we first need to prove that this set is sequentially closed. Indeed, let $\{\gamma_n\} \subset \mathcal{S}^\gamma$ (and take A_n such that $(A_n, \gamma_n) \in \mathcal{S}$) so that $\gamma_n \to \gamma$ a.s. Because $\mathcal{S}^A$ is L^0-bounded we get by the randomized Bolzano-Weierstrass Theorem 4.5.1 the existence of $A \in L^0(\mathcal{F}_t)^N$ and of randomized sequences $N_n \in \mathbb{N}(\mathcal{F}_t)$ such that $A_{N_n} \to A$ a.s. By $\sigma-$stability we have $(A_{N_n}, \gamma_{N_n}) \in \mathcal{S}$ and obviously $\gamma_{N_n} \to \gamma$ a.s. still. Because $\mathcal{S}$ is sequentially closed we get that $(A,\gamma) \in \mathcal{S}$ from which $\gamma \in \mathcal{S}^\gamma$. Therefore $\mathcal{S}^\gamma$ is indeed sequentially closed.

We proceed now to prove boundedness of $\mathcal{S}^\gamma$. Let us take $(A,\gamma) \in \mathcal{S}$ arbitrary. By applying the lower bound on c and the bound in Lemma 4.7.1 first to U^a and then to U^p, we get after recalling that $\mathcal{S}^A$ is L^0-bounded, that respectively

$$V_t(A,\gamma) \leq L_1 + U_t^p(Q_1 - \gamma\Delta R), \text{ and } V_t(A,\gamma) \leq L_2 + U_t^a(\gamma\Delta R)$$

Here $L_1, L_2 \in L^0(\mathcal{F}_t)$ and $Q_1 \in L^0(\mathcal{F}_{t+1})$. From these inequalities we get by translation invariance and time-consistency that

$$U_0^p(V_t(0,0)) \leq U_0^p(L_1 + Q_1 - \gamma\Delta R) \text{ and } U_0^a(V_t(0,0)) \leq U_0^a(L_2 + \gamma\Delta R).$$

This is valid for every $\gamma \in \mathcal{S}^\gamma$. Now take $\gamma \neq 0$, and assume that $n\gamma \in \mathcal{S}^\gamma$ for all n. Since $\gamma\Delta R \neq 0$ and because $\mathbb{E}[\gamma\Delta R] = 0$ we see that $\gamma\Delta R < 0$ and $\gamma\Delta R > 0$ on non-negligible sets. Using the same argument as in the proof of Theorem 4.6.1 we see that both $U_0^a(L_2 + \gamma\Delta R)$ and $U_0^p(L_1 + Q_1 - \gamma\Delta R)$ converge to $-\infty$ by sensitivity to large losses, contradicting the assumption that $V_t(0,0) \in dom(U_0^p) \cup dom(U_0^a)$. Therefore $\mathcal{S}^\gamma \ni \gamma \neq 0$ implies that there is an n such that $n\gamma \notin \mathcal{S}^\gamma$, from where we see that $\mathcal{S}^\gamma$ must be $L^0(\mathcal{F}_t)-$bounded.

We conclude that $\mathcal{S}$ is likewise $L^0(\mathcal{F}_t)-$bounded. ■

Now, attainability of Principal's problem under PRP:

Theorem 4.7.1 *Let us assume that $\mathbb{E}[h_{t+1}|\mathcal{F}_t] < \infty$ a.s. Further assume that $U_t^p(h_{t+1}) \in dom(U_0^p) \cup dom(U_0^a)$. Then Principal's one-step problem at time t is attained by some $(A^*, \gamma^*) \in L^0(\mathcal{F}_t)^N \times L^0(\mathcal{F}_t)^D$ and taking $\beta_t = 1$ is optimal.*

Proof. Since $V(0,0) = -c(0)\Delta t + U_t^p(h_{t+1})$ we see that $V(0,0) \in dom(U_0^p) \cup dom(U_0^a)$. The previous Lemma yields that the set $\mathcal{S} := \{V(A,\gamma) \geq V(0,0)\}$ is $L^0(\mathcal{F}_t)-$bounded. Calling $\kappa \in L^0(\mathcal{F}_t)$ the bound on the second component, we find that $|\gamma\Delta R| \leq \kappa|\Delta R| =: \tilde{r} \in L^0(\mathcal{F}_{t+1})$. Therefore, recalling that $\Gamma - \mathbb{E}[\Gamma|\mathcal{F}_t] = \gamma\Delta R$, Lemma 4.6.1 gives the desired attainability of the problem. ■

We now give an analogous result to Theorem 4.6.4, stating in the PRP case how all the one-step results can be combined together to obtain a result for dynamic contracts on the whole contracting period:

Theorem 4.7.2 *In the setting of generic functionals under PRP (see Theorem 4.7.1), and assuming the U's satisfy the usual hypotheses plus L^0-upper semicontinuity and* *sensitivity to large losses, c the super linear growth condition, and if for every time t it holds that* $U_t^p(h_{t+1}) \in dom(U_0^p) \cup dom(U_0^a)$ *plus* $\mathbb{E}(h_{t+1}|\mathcal{F}_t) < \infty$*, then the one-step problems have a solution and glueing them together yields a solution for the dynamic problem. In particular* $\beta_t = 1$ *for every t is optimal.*

Remark 4.7.1 *In light of Theorem 4.7.2 in the PRP case or Theorem 4.6.4 in the general filtration case, the claim that "it suffices to write a contract in terms of the terminal wealth" is confirmed. This is, even given the liberty to use the whole trajectory of the wealth process, the Principal can simply limit herself to more basic contracts of the forms* $\epsilon + \beta W_T$ *without losing any money. Moreover, we see that* $\beta = 1$ *is optimal. This means that the Principal essentially gives the Agent all of the money generated by the portfolio he chose, and receives in return a "derivative"* $-\epsilon$.

Remark 4.7.2 *A look at all the proofs made so far shows that* $\mathcal{F}_t$*-concavity was used essentially only to deliver* $\mathcal{F}_t$*-convexity of super-level sets (i.e. of the form* $\{f \geq a\}$*, where f is a functional) and to control the pass to sequences of convex combinations. Clearly both arguments hold if our utility functions (at time t, for every t) had more generally been* $\mathcal{F}_t$*-quasi concave, which we may define as:*

$$\forall X, Y \in L^0(\mathcal{F}_T), \forall \lambda \in L^0(\mathcal{F}_t) \cap [0,1] : U_t(\lambda X + (1-\lambda)Y) \geq \min\{U_t(X), U_t(Y)\}.$$

Indeed quasi-concavity is equivalent to convexity of all super-level sets, and in our proofs because we work with increasing limits of evaluations of functions, when performing convex combinations we would still "keep under control" the resulting sequence. For the sake of exposition and simplicity, we chose to stay in the concave framework.

4.8 Specialization to the Markovian case

Up to this point the probability space and the price process P had been rather arbitrary. In this section we will add more structure to the problem.

As an initial step we suppose price dynamics of the form:

$$\Delta P_{t+1} = P_t \left[\mu \Delta t + \sigma \Delta w_{t+1} \right]. \tag{4.8.1}$$

Here σ is a $N \times d$ matrix with linearly independent rows and $d \geq N$. We could in general assume that μ, σ are deterministic functions of time, but this does not add much value to our analysis and so we will suppose them constant. On the other hand, the process w is a $d-$dimensional martingale with finite second moments generating the filtration of the market. Such a model induces typically an arbitrage-free incomplete market. In this section we assume that both the Agent and the Principal observe w.

Notice that without loss of generality the increments of the driving process w can be assumed to be centred with uncorrelated components. We will want to have a Predictable Representation Property for w, and for that matter we know we may have to add new (adapted) components to it. We do exactly that, while keep calling without danger of confusion w to this extended process. Note that by completing σ with null rows the above dynamics still make sense for the extended process. Here the proper assumption/definition:

Assumption 20 *In the probability space $(\Omega, \mathbb{F}, \mathcal{F}, \mathbb{P})$, the filtration is generated by the realizations of the $D-$dimensional stochastic process w (that is, $\mathcal{F}_t = \sigma(\{w_0, \ldots, w_t\})$). The increments of this process are uncorrelated to each other, independent of the past, have mean zero and not-null second moments (that is $E(\Delta w_t^i \Delta w_t^k)$ is equal to 0 if $i \neq k$ and not-null otherwise), and the following Predictable Representation Property (PRP) holds:*

$$L^0(\mathcal{F}_{t+1}) = \left\{ x + Z \cdot \Delta w_{t+1} \,:\, x \in L^0(\mathcal{F}_t), Z \in [L^0(\mathcal{F}_t)]^D \right\}, \tag{4.8.2}$$

where as usual $\Delta w_{t+1} = [w_{t+1}^1 - w_t^1, \ldots, w_{t+1}^D - w_t^D]'$.

Again, if initially the $d-$dimensional w process driving the price had not enjoyed the PRP, then the previous assumption simply says that we can complete the former process obtaining something that does enjoy the PRP, while preserving "orthogonality" and not changing the informational structure of the model. At this point, it is worth making an observation about the PRP in discrete-time:

Remark 4.8.1 *Under Assumption 20, the space $L^0(\mathcal{F}_t)$ is isomorphic to $\mathbb{R}^{(D+1)^t}$. Even without "orthogonality" of the increments one can show $L^0(\mathcal{F}_t)$ to be isomorphic to a subspace of $\mathbb{R}^{(D+1)^t}$. Indeed if for instance $D = 3$ and (w^1, w^2, w^3) had the PRP, then $f \in L^0(\mathcal{F}_1) \iff f = a + b^1 \Delta w_1^1 + b^2 \Delta w_1^2 + b^3 \Delta w_1^3$ for some real constants a, b^1, b^2, b^3. In general, for every element in $L^0(\mathcal{F}_t)$ we need to specify a real number for every possible product of intertemporal increments (as well as a constant term) of the different components of w up to time t. This shows that under the PRP one could effectively reduce Principal's and Agent's problems to finite-dimensional ones, by solving such problems for every fixed random event (or tree branch) and then putting all together. However, we would need to keep track of $(D+1)^t$ real-valued functions and potentially give conditions on them for attainability of the problems. Therefore even in the case with PRP the setting and tools of finite-dimensional conditional analysis provide a more direct and elegant approach based on the original elements of the problem (as opposed to the $(D+1)^t$ real-valued functions "hidden" under a functional U_t).*

Now, some observations about the utility functionals (U^a, U^p) are in order:

Remark 4.8.2 *Evidently, the decomposition in Assumption 20 is a.s. unique. Also notice that if U is a utility functional, then $U_t(x + Z \cdot \Delta w_{t+1}) = x + U_t(Z \cdot \Delta w_{t+1})$ and therefore all the relevant information of U_t is summarized by the functional $Z \in [L^0(\mathcal{F}_t)]^D \mapsto g_t(Z) := U_t(Z \cdot \Delta w_{t+1})$, which will be called the generator of U_t. Clearly g_t inherits from U being null at the origin and concave. In the case that P may only take a finite number of values, and by the "local property," it holds $\mathbb{1}_{Z(\cdot)=z} g_t(Z)(\cdot) = \mathbb{1}_{Z(\cdot)=z} g_t(z)(\cdot)$.*

Remark 4.8.3 *Unlike in the linear expectation case, $g_t(\cdot) = U_t(\cdot \Delta w_t)$ need not be a real valued function (as opposed to $L^0(\mathcal{F}_t)$-valued) even if restricted to $\mathbb{R}^D$, because in general U_t is a non-linear operator and the increments of w are only assumed independent from the past in the linear expectation sense. This is simply to say that g decomposes into a sum or real-valued functions multiplied by products of random increments (see note 4.8.1). However, for instance when U is of the form of an optimized certainty equivalent $U_t(X) := \operatorname{ess\,sup}_{s\in\mathbb{R}}\{s - \mathbb{E}(H(s-X)|\mathcal{F}_t)\}$ then it does hold that g_t is real-valued over $\mathbb{R}^D$ and hence easier to deal with (the previous observation shows that this in turn makes U also easier to deal with). This is of course the case for entropic utility functions (taking $H(l) = \exp(l-1)$). In the converse direction, if g_t is real-valued over $\mathbb{R}^D$ and we call $\tilde{g}_t$ this restriction, and assuming that g_t is L^0-upper semicontinuous, then a decreasing approximation by step functions argument yields that $g_t(Z) = \tilde{g}_t(Z)$ a.s.*

Let us summarize some previous results to the present context, and in particular notice that recursions become backward stochastic difference equations ($BS\Delta E$s for short):

Corollary 4.8.1 *Under Assumption 20, call g^a and g^p the corresponding generators of U^a and U^p. Suppose that g^a is real-valued over $\mathbb{R}^D$ and differentiable as well as c. Agent's optimal wealth given a contract (ϵ, β) is then the unique solution of the following $BS\Delta E$ (along with a unique process Z^a):*

$$\begin{aligned} H_T &= \epsilon \\ \Delta H_{t+1} &= Z^a_t \cdot \Delta w_{t+1} - G(t, Z^a_t), \end{aligned} \tag{4.8.3}$$

where $G(t,X) = esssup_{A\in L^0(\mathcal{F}_t)}\left\{[\beta_t A\cdot\mu - c_t(A)]\Delta t + g^a_t\left(X + \beta_t A\cdot\sigma\right)\right\}$.

Therefore Principal's future utility is uniquely determined (along with a unique process $\hat{Z}^p$) by:

$$\begin{aligned} h_T &= 0 \\ \Delta h_t &= -\operatorname*{ess\,sup}_{\substack{(A,\beta,\gamma)\in(L^0)^N\times L^0\times(L^0)^D \\ [\beta\mu-\nabla c_t(A)]\Delta t+\beta\sigma\nabla g^a(\gamma)=0}} [A'\mu - c_t(A)]\,\Delta t + g^a_t(\gamma) + g^p_t\left(\hat{Z}^p_t + A\cdot\sigma - \gamma\right) \\ &\quad + \hat{Z}^p_t\cdot\Delta w_{t+1}. \end{aligned} \tag{4.8.4}$$

Proof. It is a straightforward adaptation of some arguments in Theorem 4.4.1 to the present context. In general, simply notice that if K is a translation invariant operator, and $m_s = K(m_{s+1})$, then $m_{s+1} - m_s = m_{s+1} - K(m_{s+1}) = \pi_t(m_{s+1}) - K(\pi_t(m_{s+1}))$, where $\pi_t(X) = X - \mathbb{E}[X|\mathcal{F}_t]$. By the PRP, clearly $\pi_t(X) = \text{something}\cdot\Delta w_{t+1}$, which completes the proof ■

Thus, both Principal's and Agent's problems are written as BSΔEs. In each case, we call the respective Z process *the driver*, and the $\mathcal{F}_t-$measurable term in the BSΔE (the one without the stochastic increment) the drift. Of course this could have also been done in the previous section, after the PRP assumption had been made.

In order to proceed, some more simplifications are helpful. They will render our problem a natural Markovian nature, and hence the name of this section. Take the following assumption:

Assumption 21 *Suppose that the generators g^a and g^p are Markovian, this is, $g^a, g^p : \mathbb{R}^D \to \mathbb{R}$. Also, suppose that the increments have a constant variance equal to $h := \Delta t$, this is: for all i, t: $\mathbb{E}(\Delta w_t^i) = 0$ and $\mathbb{E}\left([\Delta w_t^i]^2\right) = h$.*

In light of the previous remarks and notes, Markovianity as defined above is equivalent to having that in the decomposition of U_t into a weighed sum of products, all of the coefficients vanish except for the one associated to no-increments. This has the virtue of effectively turning the whole optimization into a finite-dimensional one, without any randomness at all. Now, some examples to better understand the issue at hand:

Example 4.8.1 *In the case of a Bernoulli walk setting, it is clear that $U_t(x\Delta w_t) = \sup_{s\in\mathbb{R}}\{s - \mathbb{E}(H(s - x\Delta w_t))\} =: g(x)$, and hence they are Markovian. In light of this and previous examples, a Bernoulli walk setting with these kind of utility functions always fulfils Assumptions 21 and 20.*

Remark 4.8.4 *From equation* (4.8.4) *it becomes apparent that under the PRP Assumption 20 and Markovianity 21, both $h_t \in \mathbb{R}$ and $\hat{Z}_t^p = 0$, for all t. Indeed, everything (the g's and c) is non-random in this case, from which it suffices to consider $(A, \beta, \gamma) \in \mathbb{R}^N \times \mathbb{R} \times \mathbb{R}^D$ in the maximization. And because $h_T = 0$, simply taking $\hat{Z}_T^p = 0$ yields $h_{T-1} \in \mathbb{R}$ (thus in $L^0(\mathcal{F}_{T-1})$). Hence, by backwards induction we see quite analogously what is claimed, provided at every step the suprema are finite. This also shows that in this case if there is an optimal contract, then (A, β, γ) is non-random. This is in accordance to a previous observation that under the Markovianity assumption the problem is essentially a deterministic one.*

Without any danger of confusion, a tuple (A, β, γ) will also be called a *contract*. From now on, the PRP-Assumption 20 and Markovian-Assumption 21 will always hold unless stated otherwise.

Of course the theorems about attainability of the (PA) problem in the previous sections are applicable in the present setting. Therefore, we next try a more direct approach which takes advantage of the fact that we can employ standard calculus techniques in the present context.

4.8.1 Computing the optimal contract: necessary conditions

A quite general existence (attainability) theory for the (PA) problems was developed under great generality in the previous sections. The basic idea was to solve a certain unrestricted problem whose well-posedness was analyzed in the general setting as well as in the case with PRP. In this section, under PRP+Markovianity assumptions and starting from the formulation (4.8.4) the attainability issue will be tackled without resorting immediately to the unconstrained variant. We will see that in fact solving this unconstrained problem is not only sufficient but necessary in a sense. For that matter, as it was just commented, the fact that all ingredients of the problem are real numbers (vectors, matrices, etc) gives us more tools to start with. Furthermore, we will be able to write down very explicitly the optimal contract.

Let us begin by deriving the First Order Conditions (FOC) for Agent's problem and Principal's problem with the aid of Lagrange multipliers:

Lemma 4.8.1 *Assume that g_t^p is once and g_t^a, c_t are twice continuously differentiable, for every t.*

Suppose a given contract (A, β, γ) is optimal for the Principal, and that for every time t the implied one-step contracts form a regular point for the corresponding constraints appearing in the drift of (4.8.4): this is, the matrix $\Big[\mu\Delta t + \sigma\nabla g_t^a(\gamma_t) \mid \beta_t\sigma\nabla^2 g_t^a(\gamma_t) \mid -\nabla^2 c_t(A_t)\Delta t\Big] \in \mathbb{R}^{N\times(N+D+1)}$ has full range (equal to N) for every t.

Then, for each t there exists a Lagrange multiplier $\lambda_t \in \mathbb{R}^N$ such that the following system admits a solution:

$$
\begin{aligned}
0 &= [\beta_t\mu - \nabla c_t(A_t)]\,\Delta t + \beta_t\sigma\nabla g_t^a(\gamma_t) \\
0 &= [\mu - \nabla c_t(A_t)]\Delta t + \sigma\nabla g^p(A_t\cdot\sigma - \gamma_t) - \nabla^2 c_t(A_t)\lambda_t\Delta t \\
0 &= \nabla g_t^a(\gamma_t) - \nabla g_t^p(A_t\cdot\sigma - \gamma_t) + \beta_t\nabla^2 g_t^a(\gamma_t)\sigma'\lambda_t \\
0 &= \lambda_t\cdot[\mu\Delta t + \sigma\nabla g_t^a(\gamma_t)].
\end{aligned}
$$

Proof. We omit the time index for simplicity. It is easy to see that the matrix

$$\Big[\mu\Delta t + \sigma\nabla g^a(\gamma) \mid \beta\sigma\nabla^2 g^a(\gamma) \mid -\nabla^2 c(A)\Delta t\Big] \in \mathbb{R}^{N\times(N+d+1)},$$

has as rows the gradients (of the components) of the constraints of the problem. Therefore, asking linear independence of these gradients equals to asking this matrix to be of full rank (equal to N). This is the standard constraint qualification (regularity) needed in order for the constrained optimization problem of the Principal to have necessary first order conditions, dependent on a multiplier λ whose existence also follow from regularity (see e.g. [Bonnans and Shapiro, 1998, Chapter 3]). Forming the Lagrangian

$$L = [A'\mu - c_t(A)]\,\Delta t + g_t^a(\gamma) + g_t^p(A\cdot\sigma - \gamma) + \lambda\cdot\{[\beta\mu - \nabla c_t(A)]\,\Delta t + \beta\sigma\nabla g^a(\gamma)\},$$

taking the partial derivatives w.r.t. $\lambda, A, \gamma, \beta$ and making them equal zero yields the system above. ■

Dropping the time index again, notice that multiplying the first equation by λ yields $\lambda\cdot\nabla c(A) = 0$. Thus multiplying the third one by $\lambda'\sigma$, the second one by λ, adding them and then multiplying by β yields:

$$\beta\lambda'[\beta\sigma\nabla^2 g^a(\gamma)\sigma' - \nabla^2 c(A)\Delta t]\lambda = 0.$$

Therefore, as soon as one searches for a $\beta > 0$ (which makes economic sense) and either c or g^a are respectively strictly convex or concave, then necessarily $\lambda = 0$. This shows that a reasonable optimal solution to the problems (meaning $\beta > 0$) must necessarily solve also the "unconstrained" problem with FOC:

$$\begin{aligned} 0 &= [\beta\mu - \nabla c_t(A)]\,\Delta t + \beta\sigma\nabla g^a(\gamma) \\ 0 &= [\mu - \nabla c_t(A)]\Delta t + \sigma\nabla g^p(\sigma' A - \gamma) \\ 0 &= \nabla g^a(\gamma) - \nabla g^p(\sigma' A - \gamma). \end{aligned}$$

This is, the Principal solves her problem without taking into account Agent's incentive compatibility constraint. We knew from previous sections, in greater generality, that solving the unconstrained problem permits (i.e. is sufficient) to construct a solution to the original constrained one. Hence these last equations show that, in the present context at least, passing through the unconstrained formulation is actually also necessary, as long as we are looking for contracts with $\beta > 0$.

As a consequence of the above equations, by summing the first two and using the third one, we get:

$$(\beta - 1)[\mu\Delta t + \sigma\nabla g^a(\gamma)] = 0.$$

Thus either $\beta = 1$ is optimal, or $\mu\Delta t + \sigma\nabla g^a(\gamma) = 0$. This last case can be called a degenerate case, since under it we easily derive that it is optimal for the Agent to exercise minimum effort: $\nabla c(A) = 0$. Since necessary conditions give a larger set of potential optimal points than the actual set of optima, we are inclined to say that this degenerate case is never optimal at all. This idea is enforced by the results in section 4.6, giving conditions under which $\beta = 1$ is indeed optimal.

As a final reassuring remark, we look at the case when both utility functionals stem from a single one, and we shall recover Theorem 4.6.3 in the present setting. Therefore, let us assume $U^a(\cdot) = \frac{1}{\gamma^a}U(\gamma^a\cdot)$ and $U^p(\cdot) = \frac{1}{\gamma^p}U(\gamma^p\cdot)$, or in terms of generators, $g^a(\cdot) = \frac{1}{\gamma^a}g(\gamma^a\cdot)$ and $g^p(\cdot) = \frac{1}{\gamma^p}g(\gamma^p\cdot)$, and suppose ∇g to be injective. Then we can always say that (A^*, γ^*, β^*) satisfies the system in Lemma 4.8.1, with $\lambda = 0$, where A^* solves

$$0 = [\mu - \nabla c_t(A^*)]\Delta t + \sigma\nabla g\left[\frac{\gamma^a\gamma^p}{\gamma^a + \gamma^p}\sigma' A^*\right], \tag{4.8.5}$$

and $\gamma^* = \frac{\gamma^p}{\gamma^a+\gamma^p}\sigma' A^*$ and $\beta^* = 1$. This is indeed in perfect harmony with Theorem 4.6.3 for the more general case, after making the proper identifications.

Let us now find the ultimate (and more explicit) structure of optimal contracts under the current assumptions:

Theorem 4.8.1 *Under the Markovianity and PRP Assumptions, the optimal contract* *is of the form (interpreted as a mapping between strategies to payments) of:*

$$A \mapsto S^A = \kappa + \sum \gamma_t^* \Delta w_{t+1} + [W_T^A - \tilde{W}_T],$$

where $W_T^A = W_0 + \sum A_t \Delta \tilde{P}_{t+1}$, $\tilde{W} = W^{A^*}$, *and* $\kappa \in \mathbb{R}$. *Here* A^* *and* γ^* *(both real vector/real valued deterministic processes) are the optimal ones for the Principal, as in Lemma 4.8.1.*

If moreover the utility functionals stem from a single one by scaling, then we can write the optimal contract as

$$A \mapsto S^A = \bar{\kappa} + \frac{\gamma^p}{\gamma^p + \gamma^a} W_T^A + \frac{\gamma^a}{\gamma^p + \gamma^a} [W_T^A - \tilde{W}_T].$$

Proof. From Corollary 4.8.1 and Theorem 4.4.1 we get:

$$\begin{aligned}
\epsilon &= R + \sum \Delta H \\
&= R + \sum [Z_t^a \Delta w + c(A_t^*)\Delta t - A_t^* \mu \Delta t - g_t^a (Z_t^a + A_t^* \sigma)] \\
&= R + \sum \left[\gamma_t^* \Delta w + c(A_t^*)\Delta t - A_t^* \Delta \tilde{P} - g_t^a(\gamma_t^*)\right] \\
&= \kappa + \sum \gamma_t^* \Delta w - \tilde{W}_T,
\end{aligned}$$

where we used γ^* coming from Lemma 4.8.1 (as well as A^*) and identified $Z^a = \gamma - A\sigma$.

This shows that the contract $A \mapsto \kappa + \sum \gamma_t^* \Delta w + W_T^A - \tilde{W}_T$ is optimal, since $\beta = 1$. Notice that $\kappa = R + \sum c(A_t^*)\Delta t - g_t^a(Z_t^a + A_t^* \sigma)$ is a constant, thanks to the Markovianity Assumption and the fact that A_t^*, γ_t^* are non-random as discussed in Remark 4.8.4.

If further the utility functionals have the same structure, i.e. they are a re-scaling of one another, we know that $\gamma_t^* = \frac{\gamma^p}{\gamma^p + \gamma^a} \sigma' A_t^*$, and plugging in this into the previous expression for the optimal contract, we conclude. ■

Remark 4.8.5 *The last expression for the optimal contract is the analogue of that obtained in Ou-Yang [2003] in continuous-time and with exponential utilities. Notice that if we drop the assumption that the utility functions are of the same form, we still get a contract, but it is potentially not adapted to the price-filtration.*

Remark 4.8.6 *The previous proof shows that* $\epsilon = \kappa + \sum \gamma_t^* \Delta w - \tilde{W}_T$ *with* $\kappa = R + \sum [c(A_t^*)\Delta t - g_t^a(\gamma_t^*)] \in \mathbb{R}$. *In case* $\gamma^* = \frac{\gamma^p}{\gamma^a + \gamma^p} \sigma' A^*$ *(e.g. utilities of the same form) then this reduces to* $\epsilon = \bar{\kappa} - \frac{\gamma^a}{\gamma^a + \gamma^p} \tilde{W}_T$ *and* $\bar{\kappa} = R + \sum [c(A_t^*)\Delta t - g_t^a(\gamma_t^*) - \gamma^p/(\gamma^a + \gamma^p) A_t^* \mu \Delta t] \in \mathbb{R}$. *In any case we see that the lump-side payment* ϵ *may be arbitrarily positive or negative as the driving random process* w *allows for. Therefore, the question of contracts with limited liability (where the contracts delivering unbounded losses to the Agent are forbidden) is certainly not covered by this analysis and is a very interesting matter on its own.*

We close this section by presenting an example where we find semi-explicit solutions. We remark that in the present, discrete-time setting, it is next to impossible to obtain fully explicit solutions, yet appropriate limits when the time step tends to zero are usually simpler.

Example 4.8.2 (1d-Bernoulli Setting, Entropic Utility) *Suppose Agent's and Principal's utility functions are respectively*

$$U_t^a(X) = -\frac{1}{\gamma^a}\log\left(\mathbb{E}\left[e^{-\gamma^a X}|\mathcal{F}_t\right]\right) \text{ and } U_t^p(X) = -\frac{1}{\gamma^p}\log\left(\mathbb{E}\left[e^{-\gamma^p X}|\mathcal{F}_t\right]\right),$$

with $\gamma^a, \gamma^p > 0$, *and that Agent's cost function is* $c(a) = \frac{a^2}{2}$. *Assume also a one dimensional market driven by a simple Bernoulli-walk setting (that is* $N = d = 1$*: one asset, one source of randomness); see example 4.7.1.*

We first observe that $g_t(x) = -\log\left(\frac{e^{\sqrt{h}x}+e^{-\sqrt{h}x}}{2}\right) = -log \circ \cosh(\sqrt{h}x)$, *from which* $\nabla g_t(x) = -\sqrt{h}\tanh(\sqrt{h}x)$. *Let us assume that the space-scale is* $\sqrt{h} = \sqrt{\Delta t}$. *From here, and manipulating* (4.8.5), *we get that the optimal action* A_t^* *at time* t *is the solution to the equation:*

$$-\frac{\gamma^a\gamma^p}{\gamma^a+\gamma^p}\sqrt{\Delta t}\sigma A_t^* = \frac{1}{2}\log\left(\frac{\sigma+(A_t^*-\mu)\sqrt{\Delta t}}{\sigma-(A_t^*-\mu)\sqrt{\Delta t}}\right).$$

As has been commented, the optimal γ_t^* *is then given by* $\frac{\gamma^p}{\gamma^a+\gamma^p}\sigma A_t^*$ *and* $\beta_t^* = 1$.

4.9 On possible extensions and conclusion

4.9.1 Possible extensions

In this last section we outline and comment on some possible extensions to the economic model considered in this chapter as well as the mathematics employed to deal with it.

First of all, we stress what has been said in Remark 4.8.6; the case of contracts with limited liability must be analyzed. This is an interesting question on its own and we can clearly see that the method we use is not well-suited, as it stands, to dealing with such a question. Indeed, a key step in dispensing with the Agent's problem (recursion) was that his utility can in principle be steered by the Principal's choice of contract, and so in particular by steering the lump-side payment ϵ, which of course in the limited liability case is restricted to a certain subset only.

A second question of interest is the pass to the limit of the model under Markovianity and the PRP, as $\Delta t \to 0$ and as a proper scaling of space is introduced (for instance, $h = \Delta t$ in example 4.7.1). So far we know heuristically that if a limit exists, then the recursions and $BS\Delta E$ that we have turn to backward stochastic differential equations (BSDE), and indeed closed form solution exists where before only implicit solutions were available in the discrete framework (see Example 4.8.2).

Another theme of interest corresponds to changing the structure of the possible contracts. In this sense, it is a challenging issue to prove that optimal contracts belong to a specific subfamily (for instance linear ones, as we have considered here) when the

Principal has access to a broader family of contracts (for instance non-linear ones, or those linear in more factors). We can at this point only describe what happens when, while keeping with linear contracts, the lump-side payment is of the form of *cash plus a fixed dynamic portfolio*, which translates as considering only $\epsilon = \tilde{\epsilon} + \sum \nu_t \Delta \tilde{P}_{t+1}$, where $\tilde{\epsilon} \in L^0(\mathcal{F}_0)$ and ν being $\mathcal{F}-$adapted are control variables for the Principal. Notice that in doing so, we are excluding most options and other financial derivatives from participating in these contracts. In this context it is not difficult to arrive similarly to optimal recursions for the Principal and the Agent, but we can see that Agent's wealth is not directly controllable by $\tilde{\epsilon}$, the way it used to be by ϵ. Nevertheless it is possible to solve the resulting one-period problems by analogous conditional analysis techniques (in fact, of the finite-dimensional kind).

Very importantly, we remark that it would be desirable to build up a more robust mathematical methodology based on conditional analysis ideas, allowing to solve a broader class of dynamic stochastic optimization problems of which the present was just an instance. We envision that the approach used here should be generalizable to a more abstract level.

As a final remark, let us point out that in this work despite the Agent and Principal playing quite different roles, the information structure was rather symmetric. A further improvement or generalization in this direction would be to consider the Agent possessing more information (i.e. a larger filtration) than the Principal as a modelling assumption. An alternative point of view would be to couple the present problem with adverse selection, in which the Principal does not know the "true type" of the Agent, but has an a-priori distribution on the possible types. As a motivation, we close this section with an example where, under asymmetry of information, the solution known in the symmetric case (whereby $\beta = 1$ is optimal) is no longer optimal.

Example 4.9.1 *For simplicity we deal with a one-period model and set $P_0 = W_0 = 0$, $\Delta t = 1$.*

Assume that beside the price $P_1 = P$ there is a random variable B correlated to it such that $\mathbb{E}[P|B] \neq 0$. Information is asymmetric in the sense that at time zero the Agent observes B whereas the Principal cannot do so. We assume for simplicity that both individuals are risk-neutral, but the Agent takes the ex-Post knowledge of B into account. Further, the cost for the Agent is given by $c(\cdot) = (\cdot)^2/2$. Finally, a contract is a couple of real numbers (ϵ, β) consisting of a payment $\epsilon + \beta AP$ in case the Agent chose A as an action.

The Agent's problem is then $\operatorname*{ess\,sup}_{A \in L^0(B)} \left\{ \epsilon + \beta A \mathbb{E}[P|B] - A^2/2 \right\}$ form which for a given contract his optimal action is $\bar{A} = \beta \mathbb{E}[P|B]$. If the Agent has a reservation utility R (which for simplicity we assume deterministic), then given β the best the Principal can do is to set $\epsilon = R + \operatorname{ess\,sup}_\omega \{\bar{A}^2/2 - \beta \bar{A} \mathbb{E}[P|B]\}$, since this gives the Agent the smallest feasible utility. We see easily that then $\epsilon = R - \frac{1}{2} \operatorname{ess\,inf}_\omega \{\beta^2 \mathbb{E}[P|B]^2\}$. Therefore Principal's problem becomes

$$\operatorname*{ess\,sup}_{\beta \in \mathbb{R}} \left\{ (\beta - \beta^2) \mathbb{E}\left[P \mathbb{E}[P|B] \right] + \frac{\beta^2}{2} \operatorname{ess\,inf} \mathbb{E}[P|B]^2 \right\},$$

the solution of which is

$$\bar{\beta} = \frac{\mathbb{E}\left[P\mathbb{E}[P|B]\right]}{2\mathbb{E}[P\mathbb{E}[P|B]] - \operatorname{ess\,inf}\mathbb{E}[P|B]^2}.$$

We thus see that $\bar{\beta} = 1$ is optimal if and only if $\mathbb{E}\left[P\mathbb{E}[P|B]\right] = \operatorname{ess\,inf}\mathbb{E}[P|B]^2$. Clearly, this is generally not the case, since P and B where quite arbitrary. Reassuringly in case that B is degenerate or independent of P, which amounts to having symmetry of information, we recover that $\bar{\beta} = 1$ is optimal.

4.9.2 Conclusions

In this work we have proved in discrete-time that the results currently know in the literature about Principal-Agent problems under moral hazard and liner contracts (with further emphasis in the setting of portfolio delegation, as in Ou-Yang [2003]) are utterly robust with respect to the modelling framework. In this sense, we have proved that it was not an accident that exponential utility functions had been successfully used for these problems in the past in order to get a rather explicit solution of the contracting problem: the point is that the usual arguments rely greatly on time-consistency and translation invariance, and so we can extend them to far more general settings through the utility functionals employed in the present work. This naturally leads to analysing conditional optimization problems in finite and infinite dimensions, for which we have established a certain approach in order to tackle them. On the interpretation side again, under this more general modelling framework we could prove that the first-best solution (whereby the Agent is surrogate to the Principal) is implementable by the Principal as had been proved in Ou-Yang [2003] in their specific framework, and we showed that the real argument behind this is the simple algebraic fact that the risk-sharing- (convolution-) type first best problem is perfectly aligned with the Agent's incentive compatibility condition as soon as in Principal's linear contracts the sensitivity-to-performance parameter β is set to 1 at every time step, in effect implying that it is optimal for the Principal to give the Agent all of the output from his decisions yet asking for a financial derivative in exchange.

Bibliography

C. Aliprantis and K. Border. *Infinite dimensional analysis*. Springer, Berlin, third edition, 2006. ISBN 978-3-540-32696-0; 3-540-32696-0. A hitchhiker's guide.

J-P. Aubin and I. Ekeland. *Applied nonlinear analysis*. Dover Publications, Inc., Mineola, NY, 2006. ISBN 0-486-45324-3. Reprint of the 1984 original.

J. Backhoff. Optimización robusta de portafolios en un mercado financiero a tiempo continuo: caso de incertidumbre no compacta o lineal. Master's thesis, Universidad de Chile.

J. Backhoff and J. Fontbona. Robust optimization in financial markets without model compactness. Preprint.

J. Backhoff and F. Silva. Some sensitivity results in stochastic optimal control: A lagrange multiplier point of view. Preprint.

P. Barrieu and N. El Karoui. Inf-convolution of risk measures and optimal risk transfer. *Finance Stoch.*, 9(2):269–298, 2005. ISSN 0949-2984. doi: 10.1007/s00780-005-0152-0. URL `http://dx.doi.org/10.1007/s00780-005-0152-0`.

F. Baudoin. Modeling anticipations on financial markets. In *Paris-Princeton Lectures on Mathematical Finance, 2002*, volume 1814 of *Lecture Notes in Math.*, pages 43–94. Springer, Berlin, 2003. doi: 10.1007/978-3-540-44859-4_2. URL `http://dx.doi.org/10.1007/978-3-540-44859-4_2`.

A. Ben-Tal, A. Ben-Israel, and M. Teboulle. Certainty equivalents and information measures: duality and extremal principles. *J. Math. Anal. Appl.*, 157(1):211–236, 1991. ISSN 0022-247X. doi: 10.1016/0022-247X(91)90145-P. URL `http://dx.doi.org/10.1016/0022-247X(91)90145-P`.

A. Bensoussan. Stochastic maximum principle for distributed parameter system. *J. Franklin Inst.*, 315:387–406, 1983.

S. Biagini and M. Frittelli. A unified framework for utility maximization problems: an Orlicz space approach. *Ann. Appl. Probab.*, 18(3):929–966, 2008. ISSN 1050-5164. doi: 10.1214/07-AAP469. URL `http://dx.doi.org/10.1214/07-AAP469`.

J.-M Bismut. Conjugate convex functions in optimal stochastic control. *J. Math. Anal. Appl.*, 44:384–404, 1973. ISSN 0022-247x.

J.-M. Bismut. Théorie probabiliste du contrôle des diffusions. *Mem. Amer. Math. Soc.*, 4(167):xiii+130, 1976a. ISSN 0065-9266.

J.-M. Bismut. Linear quadratic optimal stochastic control with random coefficients. *SIAM J. Control Optimization*, 14(3):419–444, 1976b. ISSN 0363-0129.

J. F. Bonnans. *Optimisation Continue*. Dunod, Paris, 2006.

J. F. Bonnans and A. Shapiro. Optimization problems with perturbations: a guided tour. *SIAM Rev.*, 40(2):228–264, 1998. ISSN 0036-1445. doi: 10.1137/S0036144596302644. URL `http://dx.doi.org/10.1137/S0036144596302644`.

J. F. Bonnans and A. Shapiro. *Perturbation analysis of optimization problems*. Springer Series in Operations Research. Springer-Verlag, New York, 2000. ISBN 0-387-98705-3.

J. F. Bonnans and F. J. Silva. First and second order necessary conditions for stochastic optimal control problems. *Appl. Math. Optim.*, 65(3):403–439, 2012. ISSN 0095-4616. doi: 10.1007/s00245-012-9162-4. URL `http://dx.doi.org/10.1007/s00245-012-9162-4`.

K. Borch. Reciprocal reinsurance treaties seen as a two-person co-operative game. *Skand. Aktuarietidskr.*, 1960:29–58 (1961), 1960.

G. Bordigoni, A. Matoussi, and M. Schweizer. A stochastic control approach to a robust utility maximization problem. In *Stochastic analysis and applications*, volume 2 of *Abel Symp.*, pages 125–151. Springer, Berlin, 2007. doi: 10.1007/978-3-540-70847-6_6. URL `http://dx.doi.org/10.1007/978-3-540-70847-6_6`.

C. Burgert and L. Rüschendorf. Optimal consumption strategies under model uncertainty. *Statist. Decisions*, 23(1):1–14, 2005. ISSN 0721-2631. doi: 10.1524/stnd.2005.23.1.1. URL `http://dx.doi.org/10.1524/stnd.2005.23.1.1`.

A. Cadenillas and T. Karatzas. The stochastic maximum principle for linear convex optimal control with random coefficients. *SIAM J. Control Optim.*, 33(2):590–624, 1995. ISSN 0363-0129. doi: 10.1137/S0363012992240722. URL `http://dx.doi.org/10.1137/S0363012992240722`.

A. Cadenillas, J. Cvitanić, and F. Zapatero. Optimal risk-sharing with effort and project choice. *J. Econom. Theory*, 133(1):403–440, 2007. ISSN 0022-0531. doi: 10.1016/j.jet.2005.12.007. URL `http://dx.doi.org/10.1016/j.jet.2005.12.007`.

S. Chen and J. Yong. Stochastic linear quadratic optimal control problems. *Appl. Math. Optim.*, 43:21–45, 2001.

S. Chen, X. Li, and X.Y. Zhou. Stochastic linear quadratic regulators with indefinite control weight costs. *SIAM J. Control Optim.*, 36:1685–1702 (electronic), 1998.

X. Cheng and J. Yan. A new look at the Lagrange method for continuous-time stochastic optimization. *Sci. China Math.*, 55(11):2247–2258, 2012. ISSN 1674-7283. doi: 10.1007/s11425-012-4519-3. URL `http://dx.doi.org/10.1007/s11425-012-4519-3`.

P. Cheridito and T. Li. Risk measures on Orlicz hearts. *Math. Finance*, 19(2):189–214, 2009. ISSN 0960-1627. doi: 10.1111/j.1467-9965.2009.00364.x. URL `http://dx.doi.org/10.1111/j.1467-9965.2009.00364.x`.

P. Cheridito, U. Horst, and M. Kupper. Equilibrium pricing in incomplete markets under translation invariant preferences. Preprint, a.

P. Cheridito, M. Kupper, and N. Vogelpoth. Conditional analysis on R^d. Preprint, b.

A. Cherny and M. Kupper. Divergence utilities. Preprint, 2007. URL `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1023525`.

J. Cvitanić and J. Zhang. *Contract theory in continuous-time models.* Springer Finance. Springer, Heidelberg, 2013. ISBN 978-3-642-14199-7; 978-3-642-14200-0. doi: 10.1007/978-3-642-14200-0. URL `http://dx.doi.org/10.1007/978-3-642-14200-0`.

J. Cvitanić, D. Possamaï, and N. Touzi. Moral hazard in dynamic risk management. 2014. URL `http://www.hss.caltech.edu/~cvitanic/PAPERS/dylan.pdf`.

F. Delbaen. Representing martingale measures when asset prices are continuous and bounded. *Mathematical Finance*, 2 no. 2:107–130, 1992.

F. Delbaen and W. Schachermayer. A general version of the fundamental theorem of asset pricing. *Math. Ann.*, 300(3):463–520, 1994. ISSN 0025-5831. doi: 10.1007/BF01450498. URL `http://dx.doi.org/10.1007/BF01450498`.

L. Denis and M. Kervarec. Optimal investment under model uncertainty in nondominated models. *SIAM J. Control Optim.*, 51(3):1803–1822, 2013. ISSN 0363-0129. doi: 10.1137/100782528. URL `http://dx.doi.org/10.1137/100782528`.

L. Denis and C. Martini. A theoretical framework for the pricing of contingent claims in the presence of model uncertainty. *Ann. Appl. Probab.*, 16(2):827–852, 2006. ISSN 1050-5164. doi: 10.1214/105051606000000169. URL `http://dx.doi.org/10.1214/105051606000000169`.

J. Diestel. Uniform integrability: an introduction. *Rend. Istit. Mat. Univ. Trieste*, 23 (1):41–80 (1993), 1991. ISSN 0049-4704. School on Measure Theory and Real Analysis (Grado, 1991).

D. Duffie and H.R. Richardson. Mean-variance hedging in continuous time. *Ann. Appl. Probab.*, 1(1):1–15, 1991. ISSN 1050-5164. URL http://links.jstor.org/sici?sici=1050-5164(199102)1:1<1:MHICT>2.0.CO;2-L&origin=MSN.

D. Filipović, M. Kupper, and N. Vogelpoth. Separation and duality in locally L^0-convex modules. *J. Funct. Anal.*, 256(12):3996–4029, 2009. ISSN 0022-1236. doi: 10.1016/j.jfa.2008.11.015. URL `http://dx.doi.org/10.1016/j.jfa.2008.11.015`.

D. Filipović, M. Kupper, and N. Vogelpoth. Approaches to conditional risk. *SIAM J. Financial Math.*, 3(1):402–432, 2012. ISSN 1945-497X. doi: 10.1137/090773076. URL `http://dx.doi.org/10.1137/090773076`.

H. Föllmer and A. Gundel. Robust projections in the class of martingale measures. *Illinois J. Math.*, 50(1-4):439–472 (electronic), 2006. ISSN 0019-2082. URL `http://projecteuclid.org/euclid.ijm/1258059482`.

H. Föllmer and A. Schied. *Stochastic finance*, volume 27 of *de Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, extended edition, 2004. ISBN 3-11-018346-3. doi: 10.1515/9783110212075. URL `http://dx.doi.org/10.1515/9783110212075`. An introduction in discrete time.

N. C. Framstad, B. Øksendal, and A. Sulem. Sufficient stochastic maximum principle for the optimal control of jump diffusions and applications to finance. *J. Optim. Theory Appl.*, 121(1):77–98, 2004. ISSN 0022-3239. doi: 10.1023/B:JOTA.0000026132.62934.96. URL `http://dx.doi.org/10.1023/B:JOTA.0000026132.62934.96`.

I. Gilboa and D. Schmeidler. Maxmin expected utility with nonunique prior. *J. Math. Econom.*, 18(2):141–153, 1989. ISSN 0304-4068. doi: 10.1016/0304-4068(89)90018-9. URL `http://dx.doi.org/10.1016/0304-4068(89)90018-9`.

T. Goll and L. Rüschendorf. Minimax and minimal distance martingale measures and their relationship to portfolio optimization. *Finance Stoch.*, 5(4):557–581, 2001. ISSN 0949-2984. doi: 10.1007/s007800100052. URL `http://dx.doi.org/10.1007/s007800100052`.

H. Gordon. The maximal ideal space of a ring of measurable functions. *Amer. J. Math.*, 88:827–843, 1966. ISSN 0002-9327.

A. Gundel. Robust utility maximization for complete and incomplete market models. *Finance Stoch.*, 9(2):151–176, 2005. ISSN 0949-2984. doi: 10.1007/s00780-004-0148-1. URL `http://dx.doi.org/10.1007/s00780-004-0148-1`.

A. Gundel. *Utility Maximization, f-Projections and Risk Constraints*. PhD thesis, Humboldt Universität zu Berlin, 2006.

U.G. Haussmann. *A stochastic maximum principle for optimal control of diffusions*. Pitman research notes in mathematics series. Longman, Scientific & Technical, 1986. ISBN 9780582988934. URL `http://books.google.fr/books?id=9ROoAAAAIAAJ`.

D. Hernández-Hernández and A. Schied. A control approach to robust utility maximization with logarithmic utility and time-consistent penalties. *Stochastic Process. Appl.*, 117(8):980–1000, 2007. ISSN 0304-4149. doi: 10.1016/j.spa.2006.11.005. URL `http://dx.doi.org/10.1016/j.spa.2006.11.005`.

B. Holmström and P. Milgrom. Aggregation and linearity in the provision of intertemporal incentives. *Econometrica*, 55(2):303–328, 1987. ISSN 0012-9682. doi: 10.2307/1913238. URL `http://dx.doi.org/10.2307/1913238`.

N. Ikeda and S. Watanabe. *Stochastic differential equations and diffusion processes*, volume 24 of *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam; Kodansha, Ltd., Tokyo, second edition, 1989. ISBN 0-444-87378-3.

I. Karatzas and S. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991. ISBN 0-387-97655-8. doi: 10.1007/978-1-4612-0949-2. URL `http://dx.doi.org/10.1007/978-1-4612-0949-2`.

I. Karatzas and S. Shreve. *Methods of mathematical finance*, volume 39 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, 1998. ISBN 0-387-94839-2.

C. Kardaras and G. Žitković. Stability of the utility maximization problem with random endowment in incomplete markets. *Math. Finance*, 21(2):313–333, 2011. ISSN 0960-1627. doi: 10.1111/j.1467-9965.2010.00433.x. URL `http://dx.doi.org/10.1111/j.1467-9965.2010.00433.x`.

I. Klein and W. Schachermayer. A quantitative and a dual version of the Halmos-Savage theorem with applications to mathematical finance. *Ann. Probab.*, 24(2):867–881, 1996. ISSN 0091-1798. doi: 10.1214/aop/1039639366. URL `http://dx.doi.org/10.1214/aop/1039639366`.

R. Korn and H. Kraft. Continuous-time delegated portfolio management with homogeneous expectations: Can an agency conflict be avoided? *Financial Markets and Portfolio Management, Springer*, 22:67–90, 2008.

P. Kosmol and M. Pavon. Solving optimal control problems by means of general Lagrange functionals. *Automatica J. IFAC*, 37(6):907–913, 2001. ISSN 0005-1098. doi: 10.1016/S0005-1098(01)00033-4. URL `http://dx.doi.org/10.1016/S0005-1098(01)00033-4`.

A. Kozek. Orlicz spaces of functions with values in Banach spaces. *Comment. Math. Prace Mat.*, 19(2):259–288, 1976/77.

A. Kozek. Convex integral functionals on Orlicz spaces. *Comment. Math. Prace Mat.*, 21(1):109–135, 1980. ISSN 0373-8299.

D. Kramkov and W. Schachermayer. The asymptotic elasticity of utility functions and optimal investment in incomplete markets. *Ann. Appl. Probab.*, 9(3):904–950, 1999. ISSN 1050-5164. doi: 10.1214/aoap/1029962818. URL `http://dx.doi.org/10.1214/aoap/1029962818`.

D. Kramkov and W. Schachermayer. Necessary and sufficient conditions in the problem of optimal investment in incomplete markets. *Ann. Appl. Probab.*, 13(4):1504–1516, 2003. ISSN 1050-5164. doi: 10.1214/aoap/1069786508. URL `http://dx.doi.org/10.1214/aoap/1069786508`.

D. Kramkov and M. Sîrbu. On the two-times differentiability of the value functions in the problem of optimal investment in incomplete markets. *Ann. Appl. Probab.*, 16(3):1352–1384, 2006. ISSN 1050-5164. doi: 10.1214/105051606000000259. URL `http://dx.doi.org/10.1214/105051606000000259`.

H.J. Kushner. On the stochastic maximum principle: Fixed time of control. *J. Math. Anal. Appl.*, 11:78–92, 1965. ISSN 0022-247x.

K. Larsen and G. Žitković. Stability of utility-maximization in incomplete markets. *Stochastic Process. Appl.*, 117(11):1642–1662, 2007a. ISSN 0304-4149. doi: 10.1016/j.spa.2006.10.012. URL http://dx.doi.org/10.1016/j.spa.2006.10.012.

K. Larsen and G. Žitković. Stability of utility-maximization in incomplete markets. *Stochastic Process. Appl.*, 117(11):1642–1662, 2007b. ISSN 0304-4149. doi: 10.1016/j.spa.2006.10.012. URL http://dx.doi.org/10.1016/j.spa.2006.10.012.

C. Léonard. Minimizers of energy functionals under not very integrable constraints. *J. Convex Anal.*, 10(1):63–88, 2003. ISSN 0944-6532.

C. Léonard. Minimization of entropy functionals. *J. Math. Anal. Appl.*, 346(1):183–204, 2008. ISSN 0022-247X. doi: 10.1016/j.jmaa.2008.04.048. URL http://dx.doi.org/10.1016/j.jmaa.2008.04.048.

C. Léonard. Convex minimization problems with weak constraint qualifications. *J. Convex Anal.*, 17(1):321–348, 2010. ISSN 0944-6532.

D. Li and X.Y. Zhou. Continuous-time mean-variance portfolio selection: a stochastic LQ framework. *Appl. Math. Optim.*, 42(1):19–33, 2000. ISSN 0095-4616. doi: 10.1007/s002450010003. URL http://dx.doi.org/10.1007/s002450010003.

P.D. Loewen. Parameter sensitivity in stochastic optimal control. *Stochastics*, 22(1):1–40, 1987. ISSN 0090-9491. doi: 10.1080/17442508708833465. URL http://dx.doi.org/10.1080/17442508708833465.

F. Maccheroni, M. Marinacci, and A. Rustichini. Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica*, 74(6):1447–1498, 2006. ISSN 0012-9682. doi: 10.1111/j.1468-0262.2006.00716.x. URL http://dx.doi.org/10.1111/j.1468-0262.2006.00716.x.

J.S. MacNerney. Finitely additive set functions. *Houston J. Math.*, (suppl.):iii, 1–125, 1980. ISSN 0362-1588.

D.R. Mauldin. A representation theorem for the second dual of $C[0, 1]$. *Studia Math.*, 46:197–200, 1973. ISSN 0039-3223.

H. Maurer and J. Zowe. First and second order necessary and sufficient optimality conditions for infinite-dimensional programming problems. *Math. Programming*, 16(1):98–110, 1979. ISSN 0025-5610. doi: 10.1007/BF01582096. URL http://dx.doi.org/10.1007/BF01582096.

L. Mou and J. Yong. A variational formula for stochastic controls and some applications. *Pure Appl. Math. Q.*, 3(2, Special Issue: In honor of Leon Simon. Part 1):539–567, 2007. ISSN 1558-8599. doi: 10.4310/PAMQ.2007.v3.n2.a7. URL http://dx.doi.org/10.4310/PAMQ.2007.v3.n2.a7.

J. Musielak. *Orlicz spaces and modular spaces*, volume 1034 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1983. ISBN 3-540-12706-2.

H. Nakano. *Topology of linear topological spaces*. Maruzen Co., Ltd., Tokyo, 1951.

H. Nakano. Generalized modular spaces. *Studia Math.*, 31:439–449, 1968. ISSN 0039-3223.

D. Nualart. *The Malliavin calculus and related topics*. Probability and its Applications (New York). Springer-Verlag, Berlin, second edition, 2006. ISBN 978-3-540-28328-7; 3-540-28328-5.

M. Nutz. Pathwise construction of stochastic integrals. *Electron. Commun. Probab.*, 17:no. 24, 7, 2012. ISSN 1083-589X. doi: 10.1214/ECP.v17-2099. URL `http://dx.doi.org/10.1214/ECP.v17-2099`.

H. Ou-Yang. Optimal contracts in a continuous-time delegated portfolio management problem. *Rev. Financ. Stud.*, 16-1:173–208, 2003.

S. Peng. A general stochastic maximum principle for optimal control problems. *SIAM J. Control Optim.*, 28(4):966–979, 1990. ISSN 0363-0129. doi: 10.1137/0328054. URL `http://dx.doi.org/10.1137/0328054`.

H. Pham. *Continuous-time stochastic control and optimization with financial applications*, volume 61 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2009. ISBN 978-3-540-89499-5. doi: 10.1007/978-3-540-89500-8. URL `http://dx.doi.org/10.1007/978-3-540-89500-8`.

M. Pratelli. A minimax theorem without compactness hypothesis. *Mediterr. J. Math.*, 2(1):103–112, 2005. ISSN 1660-5446. doi: 10.1007/s00009-005-0032-z. URL `http://dx.doi.org/10.1007/s00009-005-0032-z`.

P. Protter. *Stochastic integration and differential equations*, volume 21 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2005. ISBN 3-540-00313-4. Second edition. Version 2.1, Corrected third printing.

M-C. Quenez. Optimal portfolio in a multiple-priors model. In *Seminar on Stochastic Analysis, Random Fields and Applications IV*, volume 58 of *Progr. Probab.*, pages 291–321. Birkhäuser, Basel, 2004.

M. Rao and Z. Ren. *Theory of Orlicz spaces*, volume 146 of *Monographs and Textbooks in Pure and Applied Mathematics*. Marcel Dekker, Inc., New York, 1991. ISBN 0-8247-8478-2.

D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999. ISBN 3-540-64325-7.

S. Robinson. First order conditions for general nonlinear optimization. *SIAM J. Appl. Math.*, 30(4):597–607, 1976. ISSN 0036-1399.

R. T. Rockafellar. Conjugate convex functions in optimal control and the calculus of variations. *J. Math. Anal. Appl.*, 32:174–222, 1970a. ISSN 0022-247x.

R.T. Rockafellar. *Convex analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970b.

R.T. Rockafellar. *Conjugate duality and optimization*. Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1974. Lectures given at the Johns Hopkins University, Baltimore, Md., June, 1973, Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics, No. 16.

Y. Sannikov. A continuous-time version of the principal-agent problem. *Rev. Econom. Stud.*, 75(3):957–984, 2008. ISSN 0034-6527. doi: 10.1111/j.1467-937X.2008.00486.x. URL `http://dx.doi.org/10.1111/j.1467-937X.2008.00486.x`.

W. Schachermayer. *Portfolio optimization in incomplete financial markets*. Cattedra Galileiana. [Galileo Chair]. Scuola Normale Superiore, Classe di Scienze, Pisa, 2004. ISBN 88-7642-141-6.

H. Schättler and J. Sung. The first-order approach to the continuous-time principal-agent problem with exponential utility. *J. Econom. Theory*, 61(2):331–371, 1993. ISSN 0022-0531. doi: 10.1006/jeth.1993.1072. URL `http://dx.doi.org/10.1006/jeth.1993.1072`.

A. Schied. Optimal investments for robust utility functionals in complete market models. *Math. Oper. Res.*, 30(3):750–764, 2005. ISSN 0364-765X. doi: 10.1287/moor.1040.0138. URL `http://dx.doi.org/10.1287/moor.1040.0138`.

A. Schied and C.-T. Wu. Duality theory for optimal investments under model uncertainty. *Statist. Decisions*, 23(3):199–217, 2005. ISSN 0721-2631. doi: 10.1524/stnd.2005.23.3.199. URL `http://dx.doi.org/10.1524/stnd.2005.23.3.199`.

H.M. Soner, N. Touzi, and J. Zhang. Quasi-sure stochastic analysis through aggregation. *Electron. J. Probab.*, 16:no. 67, 1844–1879, 2011. ISSN 1083-6489. doi: 10.1214/EJP.v16-950. URL `http://dx.doi.org/10.1214/EJP.v16-950`.

S. Spear and S. Srivastava. On repeated moral hazard with discounting. *Rev. Econom. Stud.*, 54(4):599–617, 1987. ISSN 0034-6527. doi: 10.2307/2297484. URL `http://dx.doi.org/10.2307/2297484`.

S. Tang. General linear quadratic optimal stochastic control problems with random coefficients: linear stochastic Hamilton systems and backward stochastic Riccati equations. *SIAM J. Control Optim.*, 42(1):53–75 (electronic), 2003. ISSN 0363-0129. doi: 10.1137/S0363012901387550. URL `http://dx.doi.org/10.1137/S0363012901387550`.

N. Williams. On dynamic principal-agent problems in continuous time. 2013.

W. Wittmüss. Robust optimization of consumption with random endowment. *Stochastics*, 80(5):459–475, 2008. ISSN 1744-2508. doi: 10.1080/17442500701761347. URL `http://dx.doi.org/10.1080/17442500701761347`.

J. Yong and X.Y. Zhou. *Stochastic controls*, volume 43 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, 1999. ISBN 0-387-98723-1. Hamiltonian systems and HJB equations.

X. Y. Zhou. A unified treatment of maximum principle and dynamic programming in stochastic controls. *Stochastics Stochastics Rep.*, 36(3-4):137–161, 1991. ISSN 1045-1129.

X.Y. Zhou. Maximum principle, dynamic programming, and their connection in deterministic control. *J. Optim. Theory Appl.*, 65(2):363–373, 1990. ISSN 0022-3239. doi: 10.1007/BF01102352. URL `http://dx.doi.org/10.1007/BF01102352`.

J. Zowe and S. Kurcyusz. Regularity and stability for the mathematical programming problem in banach spaces. *Applied Mathematics and Optimization*, 5(1):49–62, 1979. ISSN 0095-4616. doi: 10.1007/BF01442543. URL `http://dx.doi.org/10.1007/BF01442543`.

Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Berlin, den 15.09.2014 Julio Daniel Backhoff